Hum 2

Vol.3 , Part 1 Jan.-March 1960

LANGUAGE and SPEECH

edited by D. B. FRY with the assistance of
Frieda Goldman-Eisler P. Denes A. C. Gimson

CONTENTS

ROBERT DRAPER LTD.

**Kerbihan House, 85 Udney Park Road,
Teddington, Middlesex, England**

Price £4 (or $11.50) per year (4 issues)

Single copies £1. 5s. (or $4)

Language and Speech

Language and Speech will be published quarterly (March, June, September and December), one volume per annum. The annual subscription is £4 (or $11.50).

Notes for Contributors

Papers are published in English only. Authors submitting material for consideration are asked to comply with the following requirements :

Typescript, which should be the original and not a carbon copy, should be double-spaced, with wide margins. The title of the paper, the author's name and initials, and the name of the institution or organization with which he is connected should be given.

Summary. A short summary should be supplied on a separate sheet and this will be printed at the beginning of the paper.

Sub-headings. Appropriate sub-headings should be inserted in the body of the paper.

References. A complete list of sources referred to in the paper should be provided on separate sheets, and in the following form :

SCRIPTURE, E. W. (1904). The Elements of Experimental Phonetics (New York).

FLETCHER, H., WEGEL, R. L. (1922). The frequency-sensitivity of normal ears. *Phys. Rev.*, 19, 553.

(Note that the titles of articles are required as well as volume and page numbers. In the text of the paper, references should be shown by giving in brackets the surname only of the author and the year of publication thus : (Scripture, 1904).

Phonetic Symbols should be restricted to those used by the International Phonetic Association and their occurrence in the text should be marked by the insertion of oblique strokes thus : /p, t, k/.

Figures should be large line drawings in Indian ink. All figure legends should be typed together on a separate sheet and numbered. The approximate position of figures and tables should be indicated in the text.

Reprints. Fifty reprints will be sent free to contributors and additional copies will be supplied at a reasonable charge.

Contributions should be addressed to the Editor :

D. B. Fry,
Department of Phonetics,
University College,
Gower Street,
London, W.C.1.

Subscriptions should be sent to the Publishers :

Robert Draper Ltd.,
Kerbihan House,
85 Udney Park Road,
Teddington,
Middlesex,
England.

CONSONANT CONFUSIONS AND THE CONSTANT RATIO RULE*

IRWIN POLLACK AND LOUIS DECKER**

Operational Applications Laboratory, Air Force Cambridge Research Center, Bedford, Massachusetts

The constant-ratio rule of Clarke was evaluated with spoken initial English consonants heard against noise: /f,h,l,r,w,y/, the cluster /hw/ and the absence of the initial consonant /#/. The average deviation between observed consonant confusions for three sets of 4 × 4 matrices and confusions predicted on the basis of the constant-ratio rule from the 8 × 8 matrix averaged about four per cent over a wide range of S/N ratios. A tentative representational structure for the selected consonants, based on the confusion analysis, is presented.

The constant-ratio rule (Clarke, 1957) summarizes the results of a confusion analysis of speech sounds in a strikingly succinct and elegant form. It states simply that the ratio among specific confusion-matrix response probabilities is invariant with the size of the matrix. Application of the rule, to date, has been encouraging (Anderson, 1959 ; Clarke, 1957 ; Clarke and Anderson, 1957 ; Egan, 1957).

The present study represents another examination of the rule, as applied to the reception of speech sounds in noise. The selected speech sounds include /f,h,l,r,w,y/, the cluster /hw/, and the absence of consonant /#/. Of these, only /f/ was studied by Miller and Nicely (1955). The selected consonants are also of interest because preliminary tests indicated that some of the consonants might be highly resistant to noise and, furthermore, until relatively recently (Lisker, 1957 ; O'Connor, *et al*, 1957) these sounds have not been examined intensively for inter-consonant confusions.

* *This is Technical Note 59-17 of the Air Force Cambridge Research Center. The research supports Project 7681, Auditory Presentation of Information, of the United States Air Force Research and Development Command Program in Human Performance. The writers are indebted to Mr. John Schjelderup for development of the stimulus-response confusion-matrix analyzer.*

** *Now with Dunlap and Associates, Washington, D.C.*

Procedure

Eight sounds /f,h,l,r,w,y/, the cluster /hw/ and the absence of a consonant /#/ were paired with /a/ (as in **father**), e.g., /fa/, /ha/, etc. Each pair was read in a carrier phrase " you will trah,————" and presented in noise to a listening crew of five university students.

Six subjects served alternately as talker and listeners. Each listener was equipped with a set of 8 response buttons connected through a scoring system which permitted the complete determination of the 8 × 8 stimulus-response confusion matrix in real time. The equipment also provided the correct alternative read by the talker following the scoring. Each of the consonants occurred equally often. Four speech-to-noise (S/N) ratios were employed : -17, -13, -9, and -5 db.

Three sets of four consonants were selected from initial set and examined at the highest three S/N ratios. These sets were /l,r,w,y/, /f,h,l,r/, and /f,h,hw,#/. The first two sets were also examined at the lowest S/N ratio. Six sets of two consonants /f,h/ /l,r/ /h,#/ /r,w/ /w,hw/ and /f,w/ were examined at a single S/N ratio of -13 db. The message sets were defined for the listeners such that the listeners' response alternatives agreed with the talker's possible message alternatives.

The number of observations associated with each alternative of each matrix was 360 observations. Thus, the total number of observations associated with each 8 × 8 4 × 4 and 2 × 2 matrix was approximately 2900, 1450 and 700 observations, respectively.

Results: Constant-ratio rule

The empirical test of the constant-ratio rule is presented in Fig. 1. Each entry in the figure represents the deviation between the obtained percentage entry in the confusion matrix and the corresponding predicted entry from the 8 × 8 matrix. (Some points in the densely packed region have been omitted). The dashed lines are drawn at deviations of ± 10 percentage points. The entries for the average correct intelligibility scores (average of the diagonals of the 4 × 4 matrices) are represented as filled points. The three different 4 × 4 matrices are coded in terms of the shape of the points.

Approximately 92 per cent of the predictions are within 10 percentage points of the obtained percentage score. One further point of interest is the systematic over-prediction of the average intelligibility scores (filled points) relative to the observed intelligibility scores.

The results of Fig. 1 are summarized in the top section of Table 1. Each entry is the mean of 16 scores, each representing the absolute deviation between an observed and a predicted score for each of the cells of a 4 × 4 matrix. The mean overall average discrepancy is only about 4 percentage points which is nearly identical with the first half vs. second half differences for entries of the 8 × 8 matrix. The bottom section of Table 1 summarizes the predictions to a 2 × 2 matrix from 8 × 8 and 4 × 4 matrices. (Some 2 × 2 combinations were obtained from more

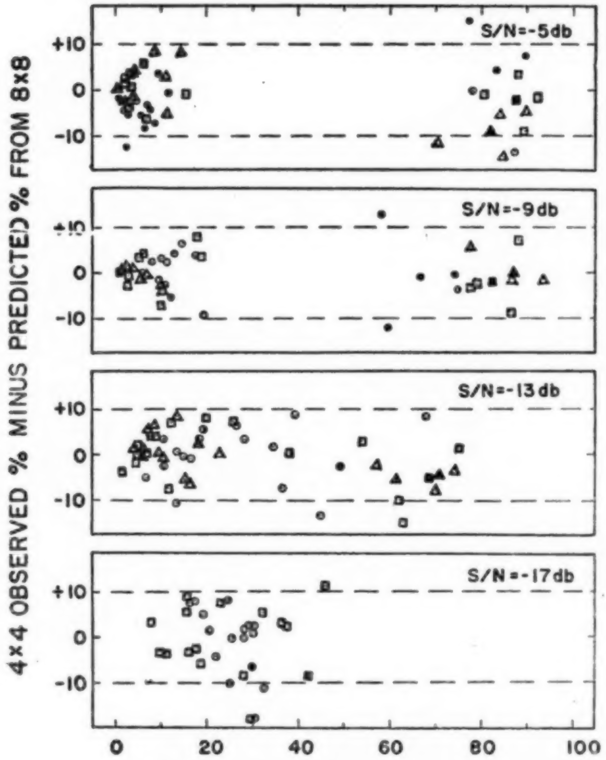


Fig. 1. Empirical test of the constant-ratio rule. The ordinate presents the average deviation for a given entry of a 4 × 4 matrix between the observed percentage score and the score predicted from the respective 8 × 8 matrix. The abscissa is the observed percentage score. Each section represents results under the indicated S/N ratio. The shape of the points is associated with three 4 × 4 matrices: /l,r,w,y/ as circles; /f,h,l,r/ as squares; and /f,h,hw,#/ as triangles. Each filled point represents the corresponding correct intelligibility score averaged over the diagonals of the matrix.

than one 4 × 4 matrix.) The average deviation between the observed and predicted score has increased to almost 8 per cent. Prediction from an 8 × 8 is not obviously poorer than from a 4 × 4 matrix for items common to the two matrices. It should be noted that the large deviations in Table 1 associated with the 2 × 2 matrices were due primarily to a systematic overestimation of the predicted scores relative to the obtained scores.

Despite the difficulties with the smaller matrices, we interpret the results in support of the constant-ratio rule.

TABLE 1

Set	8 × 8 to 4 × 4 S/N ratio in db.			
	-17	-13	-9	-5
l, r, w, y	5.5	5.2	4.5	3.7
f, h, l, r	5.9	4.7	3.4	2.8
f, h, hw, #		3.3	1.8	4.6

	S/N ratio : -13 db.	
	8 × 8 to 2 × 2	4 × 4 to 2 × 2
f, h	6.8	5.0, 5.6
f, w	15.4	
h, #	3.8	5.8
l, r	7.6	2.2, 8.1
r, w	7.4	11.3
w, hw	10.2	

Mean of the average absolute differences between the obtained and predicted entries of the confusion matrices.

TABLE 2

		Response							
S/N ratio -5 db.		f	h	l	r	w	hw	y	#
Stimulus	f	96			1	2			
	h	6	84						9
	l	1	1	76	12	5	2	2	
	r	1	1	11	57	14	5	11	
	w	1		3	5	69	15	8	
	hw	1	1	2	3	25	62	7	
	y		1	1	1	3	1	94	
	#	2	6			1			91

		Response							
S/N ratio -13 db.		f	h	l	r	w	hw	y	#
Stimulus	f	66	10	4	4	4	4	2	5
	h	14	54	4	2	2	2	1	21
	l	4	3	48	12	16	7	6	3
	r	3	3	20	27	25	9	11	1
	w	4	2	10	13	48	12	11	
	hw	9	3	4	6	26	42	10	1
	y	1	2	16	12	22	7	40	1
	#	8	20	4	3	3	2	1	60

S/N ratio -9 db.		f	h	l	r	w	hw	y	#
Stimulus	f	89	2	1	2	2	3	1	
	h	14	70	1	1	1			12
	l	4	3	63	8	12	4	5	1
	r	1	1	8	40	25	10	16	
	w	1		2	7	61	20	8	1
	hw	5	1	1	1	20	65	8	
	y	1	1	6	7	12	2	71	
	#	3	8					1	88

S/N ratio -17 db.		f	h	l	r	w	hw	y	#
Stimulus	f	28	20	12*	4	7	4	3	22
	h	8	45	14	3	7	2	6	15
	l	6	7	34	7	17	13	9	8
	r	2	7	20	18	26	8	11	8
	w	5	7	17	11	28	9	15	9
	hw	9	8	13	9	17	27	9	7
	y	3	6	17	14	23	12*	19	6
	#	13	30	9	3	4	3	6	32

*less than 11.8

Confusion matrices for 8 initial English consonants (each entry represents the percentage of responses associated with each of the stimulus items).

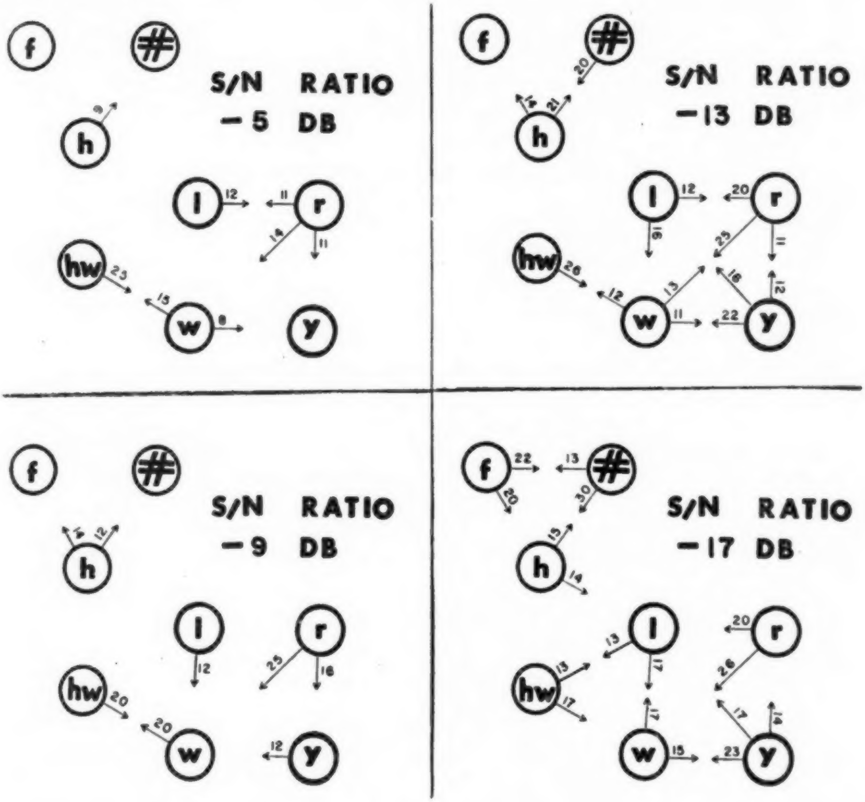


Fig. 2. Confusion vectors among eight initial English consonants. Each entry is the percentage of response confusions (direction of arrow) associated with the various stimulus consonants (origin of arrow).

Results: Pattern of constant confusions

The confusion matrices for the entire set of eight consonants are presented in Table 2. Each entry of Table 2 is the nearest rounded percentage entry of the confusion matrix.

The results of Table 2 are alternatively presented in Fig. 2 in terms of the major consonant confusion vectors. (The vectors describe the direction and magnitude of the response confusions (direction of arrow) from given stimulus consonants (origin of arrow.) For example, Fig. 2 may be read as follows: at a S/N ratio of -5 db, the response syllable /ra/ was emitted on 12 per cent of the occasions in which the stimulus syllable /la/ was read. In order to restrict Fig. 2 to the principal confusion vectors, a series of arbitrary cutoffs was imposed upon the entries of Table 2. The cutoffs associated with the several S/N ratios were adjusted in order to obtain simple structure. The cutoffs were: 7.5 per cent at a S/N ratio of -5 db; 10.1 per cent at 9 db; 10.6. per cent at -13 db; and 11.8 per cent at -17 db. (Relaxation of the cutoffs primarily serves to introduce the confusions of /r/ and /y/ with /hw/.)

The structure of Fig. 2 may be divided into three sub-groups: /l,r,w,y/ a group long regarded as possessing common characteristics, which contributes the major confusion vectors in the present study; /f,h,#/ a group which is not in evidence except at the most unfavourable S/N ratios; and, finally, the /w,hw/ grouping. We may also note a marked asymmetry of the confusion vectors; in particular, the /hw,w/ confusion was nearly twice that of the /w,hw/ confusion. The linguistic specification of these asymmetric confusions, however, is complicated by strong response biases in favour of some response alternatives over others. For example, as a result of the greater response frequency of /w/ relative to /hw/, the strong /l,r,y/ confusions with /w/ are sharply attenuated with respect to /hw/.

We finally note the broadening of the confusion matrices with more unfavourable S/N ratios. The confusions obtained at more favourable S/N ratios tend to persist to the more unfavourable S/N ratios, while additional consonant confusions are added at the more unfavourable S/N ratios.

Conclusions

The constant-ratio rule of Clarke was examined and was supported for the set of eight initial English consonants /f,h,l,r,w,y,hw,#/. Inter-confusion analysis suggests the following representational structure for these consonants: /l,r,w,y/ over a wide range of S/N ratios; /f,h,#/ at the most unfavourable S/N ratios; and a marked asymmetric confusion between /w,hw/.

References

Anderson, C. D. (1959). The Constant-Ratio Rule as a Predictor of Confusions among Visual Stimuli of Brief Exposure Duration. Hearing and Communication Laboratory, Indiana University, Bloomington, Indiana, AFCRC TN 58-60, ASTIA AD-160 706.

Clarke, F. R. (1957). Constant-ratio rule of confusion matrices in speech communications. *J. acoust. Soc. Amer.*, **29**, 715.

Clarke, F. R. and Anderson, C. D. (1957). Further test of the constant-ratio rule in speech communication. *J. acoust. Soc. Amer.*, **29**, 1318.

Egan, J. P. (1957). Monitoring task in speech communication. *J. acoust. Soc. Amer.*, **29**, 482.

Lisker, L. (1957). Minimal cues for separating /w,r,l,y/ in intervocalic position. *Word*, **13**, 256.

Miller, G. A. and Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *J. acoust. Soc. Amer.*, **27**, 338.

O'Connor, J. D., Gerstman, L. J., Liberman, A. M., Delattre, P. C. and Cooper, F. S. (1957). Acoustic cues for the perception of initial /w,j,r,l/ in English. *Word*, **13**, 24.

STATISTICAL APPROXIMATIONS TO ENGLISH AND FRENCH

ANNE TAYLOR AND NEVILLE MORAY*
Institute of Experimental Psychology, Oxford

This paper presents samples of English and French prose representing approximations to normal English and French. The range covered in English is from 2nd order to 16th order approximations and in French from 1st order to 8th order.

INTRODUCTION

In his original papers on the development of information theory, Shannon (1949) discussed the formation of approximations to words and languages in terms of choosing words consecutively in the light of varying amounts of context. Some such lists were compiled by Miller and Selfridge (1950), and also by Deese (personal communication, 1958). But all these samples have been short, and none of them have gone beyond an 8th order approximation to normal prose. In 1958 the authors published a report on dichotic listening in which they made use of some new lists which they had compiled which were greater in length and extended up to 16th order approximations. More recently we have compiled a set of French approximations, and the two sets of approximations, French and English, are presented in this paper.

METHOD

The method used in the construction of the lists is as follows. A second order list is one in which the subject adds the next word to a sentence of which he can see only a single preceding word. A 12th order list is one in which a subject adds the next word to a sentence of which he can see the preceding eleven words, and so on. By way of example, let us consider a 5th order approximation. In this a subject was shown four words of a sentence written underneath each other on a strip of paper, and was told that they were part of a sentence. He was asked to add the next word in such a way as to " try to keep the sentence going ". When he had done this the top of the paper was folded over so that still only four words were visible, and the next subject was shown these four. This process was repeated until lists of the desired length had been obtained. Thus lists of prose were obtained in each of which the effect of context differed by a known amount from that in other lists, and which can be thought of as containing each a different amount of information whose relative, though not absolute, value is related to the order and can be assessed. For a discussion of this point see Moray and Taylor (1958).

* *Now at the Department of Psychology, Hull University.*

One point which should be emphasised is that there appear to be differences of structure between these and the American lists referred to above. One source of these differences appears to be the exact method used and the instructions given when they are being compiled, and comparisons of experiments using different samples of these approximations should therefore be treated with circumspection.

Material

Statistical Approximations to English

2nd Order

The camera shop and boyhood friend from fish and screamed loudly men only when seen again and then it was jumping in the tree is idiotic idea of almost there is cabbage a horse which was not always be the set works every evening is heaviest with bovine eyes looking sideways chewing toffee stuck round about philosophy many an every cat eats his bicycle chain mail photograph of of my uncle can travel in case you must go to dance the silly fool who may yet I think straight on the rat tail of many people are incorrigible mariners sailing along (100) the cat sat happily married in the essential for fun is young children are soon hoping desperately ill will you . . .

4th Order

The boy went to eat fish on plates is only washed up on Sunday evening people sing so loudly in his examination for his certificate papers are obtained by adding lots of baking powder until consistency is right through the middle hole often might go out thoughtfully making his bed badly needs eating so eat lots of cheese in a dish was brought forward next month to see if any people can revolt me for when it exploded it killed him without blood flowing fast and furious pace of progress varies with experience but without a single thought of danger (100) came out of England to become richer quickly so they were all feeling well and happy now for ever they sing sweetly just . . .

6th Order

I have a few little facts here to test need lots of time for studying and praying for guidance in living according to common ideas as illustrated by the painting which is hanging on our most precious line and most valuable accessory to the little guidebook of England choose small villages where everyone is concerned about the houseparty at home was successful beyond words without any help to interpret the messages correctly without having made transcriptions of speeches given by one man standing as still as any rock upon a hillside where there were several people jumping about and screaming (100) shrilly giving vent to suppressed feelings . . .

8th Order

In the early days following my first enquiry I thought of asking if he could hope to win everlasting renown by writing a complete set of confused accounts which do not conform to any regulations known by any person who is able and trustworthy can decide what to make with it after you have sorted among my papers give me some shocks which could paralyse instead of cure him so that he recovered from maladies contagious and infectious which might cause much suffering on account of dust collecting where it ought not to because of the plague which took place (100) in September showers and fresh breezes blowing . . .

12th Order

Down the road there came a cart with two men riding behind the horse which trotted gently along the road toward home was blocked by an avalanche so as to prevent catastrophe from becoming too much to bear left him calm and resolute knowing without doubt that heaven was full of angels which are all different from all men or women in shape and size and also in their manner befitting such persons of rank and wealth and honesty and integrity who are not affected by prejudice or bias in judging the relative values of principles and ethics however great (100) and painful . . .

16th Order

The very next day they went for a picnic in the woods which they thought would be suitable for them to visit although they saw bulls in the field which might attack the girls savagely and so cause them untold fear no harm whatever will come providing that they flee all possible dangers arising from the long since completed series of adventures they began last week on work which took them far too far away to another village green on which they played at bowls like Drake on Plymouth Hoe before the appearance of the first galleon its sails white against (100) the . . .

Statistical Approximations to French

1st Order

Le ses béantes habitude murmura de une aussi pour me deux tête il composée de était que décrivait place il de seule belle se posa les le aux qui après sa comme pas faut parfait vient danger ou très avait si période dures de groupe mis bien ce chante le régaler rigoureuse bouquet vide ensuite regardait fait son pour vers affectueusement franchir noble à maison garde de il se le qui tard était il a pas plaisir mon ferait portière hommes sans douta ferait paternelle et cependant et j'ai coucher croix le il en plaque était salon révolu là . . .

2nd Order

Dans le journal du pôle nord est très intéressant et du fromage du sel est très gentille petite fleur jaune et alors le perroquet vert foncé comme le premier ministre d'état avait oublié son oreille rose tendre jeunesse dorée du pont d'Avignon en dormant paisiblement païssant l'herbe verte était petit mais quand tu penses profondément émue comme jamais entendu la dernière fois que tu juges avec le bourreau prit son fils de nombreux ennuis financiers sont désagréables parce que des chapeaux verts se balance les jolies paquerettes fleurissent le chandail bleu roi fainéant malgré lui et enfin bref il coulait . . .

4th Order

Je disais que tu es plus gentille que moi surtout ennuyeuse dans une pièce rose peinte par un artiste que nous aimons plus rire que quand on entend la retraite de vieillesse dans la plupart des jeunes gens élégants dont les manières étaient vulgaires malgré l'effort soutenu par Monsieur le Maire présidait l'assemblée réunie au bord du lac il contemplait le paysage ou nous cueillons des bouquets bariolés ornent tous les murs salis par les atteintes de la maladie affreuse qui a terrorisait en agitant les branches des arbres fleuris tout ensoleillés lui étaient agréables pour lui plaire car il . . .

5th Order

Le train s'est arreté pour siffler avec un petit air mutin qui me semblait drôle surtout quand il veut se reposer sur un divan confortable de style moderne en dix couleurs chatoyantes qui semblaient rayonner de bonheur depuis le jour ou sa mère avait élevé un agneau avec amour et voulait l'embrasser en rêve pendant que tu veillais l'enfant pendant qu'il dormait dans l'extase avec la belle jeune fille qui rêve de devenir un prince charmant pour danser la polka devant une salle pleine qui était infestée de vipères grouillantes dont les yeux brillaient de malice quand les garçons courtisent . . .

6th Order

La campagne est très jolie et je l'admire beaucoup d'autant plus que je venais de dire n'était jamais méchant que si on lui pardonne il sera sage et tout content d'obéir à sa mère malgré le chagrin qu'elle ressentait elle survécut vingt ans et mourut sans dire mot aux amis qui essayaient leur nouveau costume rose devant le grand ciel nuageux tourmenté de tempête hurlante dans la nuit réveilla subitement son mari en chantant et après le déjeuner elle se repose à loisir sans ombre de souci heureux et fier encore que ce monsieur lui semble beau malgré son nez . . .

8th Order

Le jeune homme est allé faire des découvertes dans des pays sauvages habités par des animaux féroces et dangereux qui heureusement mangent tous aliments divers que nous leur donnons sont nourissants et appétissants sauf si le grand mécontentement survient et soulève une désastreuse révolte comme celle de l'année précédente quand l'hiver commençait avec le vent qui sifflait dans la mâture complexe du grand bateau qui se balance mollement sur l'eau moirée recouverte de nénuphars blancs qui flottaient sur la rivière ressemblaient à quelque fantôme effrayant circulant autour de la maison en hâte sur son grand cheval éperonné et couvert de . . .

References

Miller, G. A. and Selfridge, J. A. (1950). Verbal context and the recall of meaningful material. *Amer. J. Psychol.*, **63**, 176.

Moray, N. and Taylor, A. (1958). The effect of redundancy in shadowing one of two dichotic messages. *Language and Speech*, **1**, 102.

Shannon, C. and Weaver, W. (1949). The Mathematical Theory of Communication. (Urbana, Illinois).

TIME FACTORS IN PERCEPTION OF A DOUBLE CONSONANT

J. M. PICKETT* AND LOUIS R. DECKER**
Operational Applications Laboratory, Air Force Cambridge Research Center, Bedford, Massachusetts

A listening experiment was carried out to examine the perceptual distinction between a single stop consonant, /p/, and its double counterpart, /p-p/. The joint effects of two time factors are studied: (1) the duration of /p/-closure (silence) and (2) the rate of utterance of the surrounding test sentence. The test sentence, *He was the topic of the year*, was recorded on tape and then, in a number of recorded copies, the duration of the /p/-closure was altered by inserting or removing tape. A group of listeners judged each of the altered sentences to be either *He was the topic of the year* or *He was the top pick of the year*. Effects of ten closure durations (60, 100, 150, 200, 250, 300, 350, 400, 500, and 585 msec.) are studied in various combinations with five rates of utterance (2, 3, 4, 6, and 8 syllables per second). A threshold closure duration is defined to be the duration at which 60% of the judgments were *topic*. As the rate increased from 2 to 8 syllables per second, the threshold closure duration decreased from 320 to 140 msec. and at a progressively declining rate. This function of threshold closure duration *vs.* rate of utterance is found to be approximately parallel to the relation, for the unaltered sentences, between actual closure duration and rate of utterance.

Durational cues to the various perceptual distinctions among consonants have recently been studied by Denes (1955), Lisker (1957), Gerstman (1956, 1957), and Liberman, *et al.* (1958). A further, and perhaps more obvious, durational cue may be associated with the spoken English distinction between an intervocalic single consonant and its double counterpart. This distinction, for example, differentiates the word *topic* from the phrase *top pick*. Stetson (1951) made extensive studies of the articulatory differentiation of single and double consonants. He measured closure durations for double and single stop consonants. Distinct double articulations (but with no release between the two consonants) had a median closure duration from onset to release of 200 msec. This type of double articulation normally occurred at utterance rates of 2 syllables per second (sps). Single consonant articulations had a modal closure duration of 100 to 140 msec. and generally occurred at syllable rates of 3.5 to 4 sps. (Stetson, 1951, pp. 60-74, especially p. 63 and Figs. 41, 43, and 51).

In the present experiment the perceptual distinction of a single from a double stop is briefly examined as a function of closure duration and rate of utterance. The general method was to record on tape a sentence containing a single stop closure which

† *This is Technical Note 59-13 of the Air Force Cambridge Research Center on research under Project 7681, Auditory Presentation of Information, of the Air Research and Development Command.*

* *Now at Naval Research Laboratory, Washington, D.C.*

** *Now at Dunlap and Associates, Washington, D.C.*

could then be artificially lengthened so as to cause the stop to sound double and the sentence to change in meaning. The distinction between *topic* and *top pick* was chosen for study. The word *topic* was spoken in the sentences, *He was the topic of the year* and *He was the topic for discussion.*

TESTS WITH NORMAL RATE OF UTTERANCE

Two male talkers (the authors) recorded the sentences to be manipulated. It was desired to have examples of the sentences which would be relatively ambiguous as to whether *topic* or *top pick* had been intended by the talkers. Thus, they practised and recorded repeatedly so as to provide final examples having three characteristics: (1) a relatively standard rate of utterance, (2) a /p/-closure of 150 msec., and (3) a relatively even syllable stress throughout the sentence. Under these conditions, artificial shortening of the /p/-closure caused a definite impression of *topic* upon playback ; and conversely, lengthening the closure gave definite judgments of *top pick*.

The /a/ of *topic* was 1 to 2 db more intense than the /ɪ/. The average syllable rate for the original sentences was about 4 sps. One talker recorded both of the sentences, while the other talker recorded only one. Tape speed was 15 inches per second.[1]

Each of the three recorded sentences was copied ten times and then the /p/-closure of each copy was lengthened or shortened to provide a set of ten sentences with ten different closure durations as follows: 60, 100, 150, 200, 250, 300, 350, 400, 500, and 585 msec. The closure was located by listening to the tape output as the tape was pulled slowly by hand across a playback head. Shortening was accomplished by removing tape from the middle of the closure. Lengthening was accomplished by inserting blank tape at the centre of the closure. One millisecond was equivalent to about 1/64 inch of tape. Splices could be controlled well within 1/16 in., or 4 msec. The splices were silent. The new closure durations were found to be the intended durations within 2 msec.

Test tapes of each talker were recorded by copying, in a random order, ten of each of the ten sentences with different closure durations.

Various crews from a pool of 15 college undergraduates, plus one of the authors, served as listeners. A constant crew was maintained for a given test tape. The undergraduates were naive phonetically but they were trained observers who also served in other psychophysical experiments.

The listeners were provided with written copies of the two alternative sentences for a test, and then given the following instructions: *These two sentences have been*

[1] *Recording, copying, and playback of the sentences were carried out with an RCA 88 microphone, Magnecorder Series PT-6 recorders, and an 8-in. Altec-Lansing loudspeaker mounted in a 1.5 cu.-ft. enclosure. The systems had reasonably uniform frequency response over the range 100 to 7000 cps. The recording and playback room was a furnished office 7 × 20 × 10 ft. with a hard-surfaced interior. The microphone was 6 in. from the mouth of the talker and turned 70° off zero incidence.*

spoken in a random order on this test tape. Listen to each sentence and then mark on your score sheet which sentence the talker was saying. The first eight sentences of the test tape were then played back for practice. The experimenter answered any questions about the test procedure and then played the entire test tape of 100 sentences.

The results of the tests with normal rate of utterance are shown in Fig. 1. The ordinate of Fig. 1 is the percentage of *single* ("topic") judgments. The abscissa is the closure duration of /p/ in msec. The top graph shows results for the sentence *He was the topic of the year* with separate curves for each talker. The bottom graph shows the results for *He was the topic for discussion,* spoken by only one talker.

In general, most /p/-closures shorter than 150 msec. were judged as single consonants and closures longer than 250 msec. were judged as double consonants. This result is quite consistent with Stetson's articulatory measures referred to in the introductory discussion.

Minor features of the results are: (1) the two talkers give very similar relations between closure duration and *single* consonant judgments, and (2) there is a bias toward making *single* consonant judgments. Even at the longest closure durations 5% of the judgments are *single.* This probably reflects a bias in the time and stress patterns of utterance which, despite our efforts to prevent it, apparently favoured single consonant judgments.[2]

Tests with various rates of utterance

Procedures for these tests were identical with those using the normal rate, except as noted below.

Each talker practised and then recorded *He was the topic of the year* at five different rates of utterance spanning a range from very slow to very rapid utterance. The rate was paced with the aid of the sweep hand of a 6-rpm stopwatch. The talker watched the moving hand and spoke the sentence at a rate to occupy about the time desired. The desired times were calculated on the basis of eight syllables for the sentence. Rates of 2, 3, 4, 6, and 8 sps. were chosen for testing. It was estimated that the rates were attained within an error of 10%. New recordings of each talker were then prepared by splicing as follows. For each of the two slow rates, six /p/-durations were prepared: 300, 400, 500, 600, 700, and 800 msec. For each of the three remaining rates, five /p/-durations were prepared: 50, 100, 200, 300, and 400 msec. Thus there were 27 combinations of rate and closure.

A test tape of 108 sentences was made for each talker, consisting of 4 repetitions in

[2] *Professor Meyer-Eppler suggested that the formant transitions to a /p/-locus for the /a/ of* topic *would differ from that for* top pick. *He kindly recorded in his laboratory spectrograms of Talker P speaking the two isolated utterances and also the test sentence in the manner of the experiment. The total amount of transition was the same for the two cases, approximately 200 cps for F_2 and 250 cps for F_3. However, the F_3 transition of /a/ in* topic *was complete about half way through the vowel, whereas for* top pick, *the F_3 transition occurred throughout the vowel. Vowel durations were not significantly different for /a/ but /ɪ/ was about 27% shorter in* topic *than in* top pick.

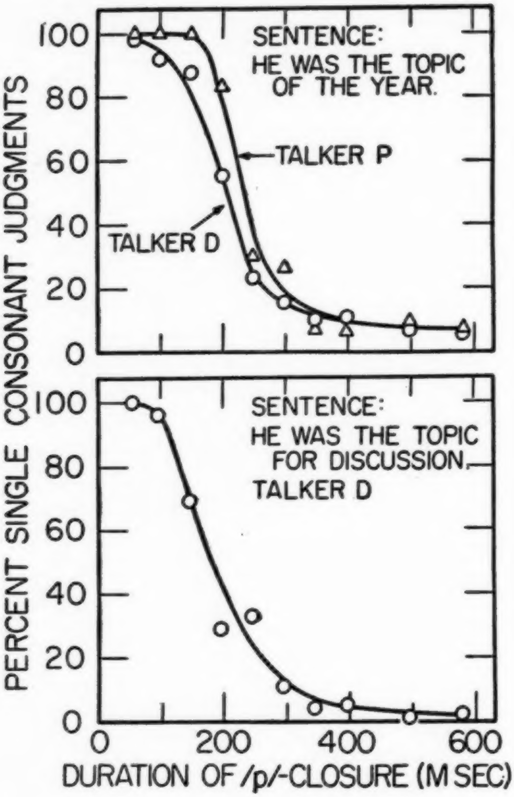


Fig. 1. Perception of a single /p/ *vs.* a double /p-p/ as a function of the closure duration of the /p/. The ordinates show the percentage of listeners' judgments which were single /p/. The abscissa of each graph is the closure duration of the /p/ in the word *topic*. When the closure is 250 msec. or longer most of the *topic*'s are judged as *top pick*. Data are shown for two test sentences and two talkers. In the top graph each point for Talker D is the result of 150 judgments: 15 listeners x 10 judgments; for Talker P, 60 judgments (6 listeners x 10 judgments) determined each point. In the bottom graph each point represents 80 judgments: 8 listeners x 10 judgments. Rate of utterance about 4 syllables per second.

a random order of each of the 27 combinations of rate and /p/-closure. The test tapes were played back to listeners to be judged as before, with the additional instruction that the talker now "*speaks in different ways.*"

The results of the tests with various rates of utterance are shown in Fig. 2. The ordinate of Fig. 2 is the percentage of *single* consonant judgments. The abscissa is the duration of /p/-closure in msec. The parameter is the average rate of utterance in

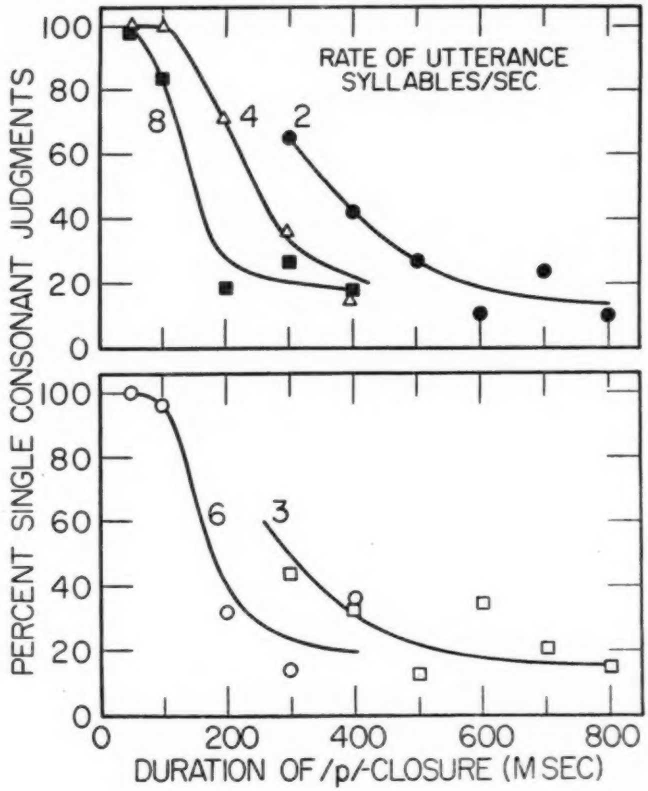


Fig. 2. Single /p/ perception *vs.* duration of /p/-closure at various rates of utterance. The ordinates show the percentage of listeners' judgments which were single /p/. The abscissa of each graph is the closure duration of the /p/ in the test sentence, *He was the topic of the year.* The parameter of each graph is the average rate of utterance within an error of about 10%. As rate of utterance decreases from left to right, longer /p/-closures are required to shift the judgments to double /p-p/. Each point is the result of 80 judgments: 2 talkers x 10 listeners x 4 judgments.

syllables per second. Results were pooled from both talkers.

The results show an interaction between rate of utterance, the perception of the consonant, and its duration. As rate becomes slower, a longer duration is necessary for the consonant to be heard as double. With short durations and fast rates, the interaction seems less marked. The curve for " normal " rate (4 sps.) is very similar to the " normal " curves of the top graph of Fig. 1.

A bias toward making the *single* judgment is again present in the results. Even the longest closures are judged *single* 10-30% of the time. Apparently, as appeared

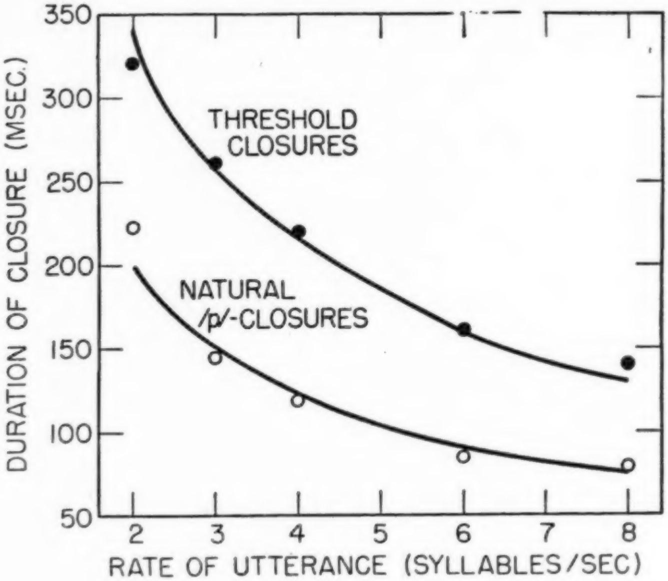


Fig. 3. Showing the similarity of perceived threshold /p/-closure and natural /p/-closure. Threshold /p/-closure (top curve) is the duration of /p/-closure that is judged single-/p/ 60% of the time, taken from the curves of Fig. 2 at each rate of utterance. Natural /p/-closure (bottom curve) is the mean duration of /p/-closures for the two talkers which occurred in the original test utterances and which were then altered in duration for the tests of Fig. 2. The abscissa is the average rate of utterance. The curves through the points are reciprocal functions described in Footnote 4.

in the first tests, we were not completely successful in achieving a style of utterance having no bias toward the *single* perception.

Many students of speech perception have postulated or demonstrated a close relation between the domains of perception and articulation.[3] Therefore, it would be interesting to compare a perceptually determined relation between consonant closure and rate of utterance with the same relation determined by articulatory measures. To the author's knowledge there were no published data on the articulatory relation. However, we had already measured the durations of our own single /p/-closures in order to know

[3] *See the following: Stetson, 1951, Appendix V; Jakobson, Halle, and Fant,* Preliminaries to Speech Analysis, *Tech. Report 13, 1952, Acoustics Laboratory, Massachusetts Institute of Technology; Liberman, Delattre, and Cooper,* Amer. J. Psychol., *1952,* 65, *497; Miller and Nicely,* J. acoust. Soc. Amer., *1955,* 27, *338; Liberman, A. M.,* J. acoust. Soc. Amer., *1957,* 29, *117; Gerstman (1957), p. 113-117; Lisker (1957), p. 47, lines 12-16; Pickett,* Language and Speech, *1958,* 1, *288.*

how much to adjust them for the experimental tests. The natural /p/-closures were available at each rate, one for each talker; these were averaged over talkers and plotted against rate to give the lower set of points in Fig. 3. It should be borne in mind that these data were generated by the same utterances used for our perceptual tests. The upper set of points in Fig. 3 represents a perceptual relation between threshold /p/-closure and rate. The perceptually determined values of threshold closure are those at which 60% of the judgments are *single* on the curves of Fig. 2. The trends of the articulatory and perceptual data are very similar, thus providing another bit of evidence for the idea that the acoustic patterns of speech are interpreted by the listener in terms of articulation.[4]

[4] *In Fig. 3 the duration of natural /p/-closure appears to be approaching a lower limit of about 75 msec. as rate increases. This is probably due to the approach to a physiological upper limit on rate of utterance. The 75-msec. value and the apparent upper limit of slightly more than 8 sps. correspond well with the rate limit and associated closure durations reported by Hudgins and Stetson (1937) from a larger set of articulatory measurements to determine the maximum rate. The two curves drawn in Fig. 3 are portions of the functions* $T = \frac{5.6}{R^{0.7}}$ *(top curve) and* $T = \frac{3.3}{R^{0.7}}$*, where* T *is the duration of /p/-closure in units of 1/10 sec. and* R *is the rate of utterance in syllables per second. These curves are cases of the function* $T = \frac{k}{R^a}$*, which, when* $a = 1$*, describes a simple reciprocal relation between the duration of a speech sound and rate of utterance. When, as in Fig. 3,* $0 < a < 1$*, the rate of decrease of* T *with increase of* R *is not as rapid as the rate of decrease of* T *when* $a > 1$*. The slower decrease may in general hold for consonant durations while the faster decrease may be typical for some vowels.*

References

Denes, P. (1955). Effect of duration on the perception of voicing. *J. acoust. Soc. Amer.*, **27**, 761.

Gerstman, L. J. (1956). Noise duration as a cue for distinguishing among fricative, affricate and stop consonants. Abstract in *J. acoust. Soc. Amer.*, **28**, 160.

Gerstman, L. J. (1957). Perceptual dimensions for the fricative portions of certain speech sounds. New York University Dissertation (University Microfilms, Inc., No. 24, 867).

Hudgins, C. V. and Stetson, R. H. (1937). Relative speed of articulatory movements. *Arch. néerl. Phon. expér.*, **13**, 85.

Liberman, A. M., Delattre, P. C., and Cooper, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position, *Language and Speech*, **1**, 153.

Lisker, L. (1957). Closure duration and the intervocalic voiced-voiceless distinction in English. *Language*, **33**, 42.

Stetson, R. H. (1951). Motor Phonetics (Amsterdam).

WORD SCALES FOR DEGREES OF OPINION

STUART CARTER DODD AND THOMAS R. GERBRICK

Washington Public Opinion Laboratory, University of Washington, Seattle

This paper describes experiments relating to the choice of words and phrases for use in questionnaires and public opinion polls. It is concerned particularly with expressions of degree or intensity of belief or opinion, and of temporal frequency.

Sets of phrases were presented to groups of subjects in random order, in serial order, and in context and the subjects were asked to place each item on a nine-point scale. From the results, the mean scale position and the ambiguity of each item were calculated ; these data were combined with an index of the length of each item (in words per syllable) and the frequency of its occurrence in the language generally to provide criteria for the choice of suitable word-scales. A number of recommended sets of phrases are given, together with their appropriate scale ratings, positions and ranges.

ASSUMPTIONS AND INTENT OF THIS STUDY

Pollers often follow up a pro-con content question by asking about its intensity, " How strongly do you feel . . . (what you have just said) . . . ? " (Cantril, 1947). It is substantively assumed here—though much more experimental evidence is needed —that the more intensely an opinion is expressed, the more its expressor will tend to : (a) resist changing it, (b) act on it, including expressing it to others, and so (c) strengthen the opinion in a potentially like-minded group of hearers.

These three assumptions imply that people with least intense or least strongly held opinions will tend to be those who are most modifiable, inactive, and non-propagating.

It is methodologically assumed here that more exact measurement of asserted intensity of feeling and other kinds of opinion in degrees will improve the reliability, validity, predictivity, and generality of intensity polling. Let " reliability " be operationally defined as degree of agreement on re-observing under like conditions—as measurable by a correlation coefficient between a poll and a re-poll. This reliability correlation is always attenuated or reduced by errors of measurement which largely mean uncorrelated variation in the poll's measurement of the respondent's opinion, whatever it is.

Let " validity " be similarly defined as the correlation between an indicator of something and an accepted criterion of that something. In polls the validity of an intensity index thus is the correlation between the polled assertion of intensity and the life behaviour elsewhere which shows that intensity of feeling. Validity in a poll is thus the degree of agreement between speech and action, between poll response and corresponding life responses. The amounts by which this validity correlation is lowered by unreliability in certain situations has been mathematically derived and can be read from a table (Edgerton and Yoops, 1928). From this limited evidence it is assumed here that more exact measurement of intensity will tend to improve its validity.

Let " predictivity " be similarly defined by the correlation between any intensity variable and any other variable to be predicted. Call these the predicter intensity and the predictand. Then it can be shown that uncorrelated errors in measurement will lower this predictivity because such errors in the variables always attenuate their correlations (Guilford, 1936).

Let " generality " be defined here as wide applicability of these intensity scales to diverse content opinions. The words like " more ", " less ", " none " can be combined with a great range of opinion content. They are not as limited to specific situations as the usual attitude scales whose items are sentences. Consequently, this generality of context is also expected to be generalizable to different (English-speaking) populations in different places and periods.

Therefore, since pollers want more reliable, valid, predictive, and generalizable polls, more exact measurement of the intensity variable (among other predicters) may be expected to contribute something (Dodd, 1942, Ch.7).

It is further assumed that errors, or chance fluctuations, will be reduced in proportion as questions in a questionnaire are well scaled. A good scale will be (a) a unidimensional scale, i.e., one showing degrees of **one kind** of entity, (b) a cardinal scale, i.e., having equal units or equal intervals, (c) a range-specified scale, i.e., having its origin or zero point and its limiting points known.[1]

An operational definition of a good opinion scale is the Kilpatrick-Edwards (1948) technique of scale construction which can combine the features of the Likert scale, the Cornell technique for ordinal scales, and the Thurstone equal-interval technique for more cardinal scales.

The present study, based on the assumptions above, is intended to provide pollers with better scales for intensity and other degrees of opinion. It reports intensity " scalettes ", or sets of scaled words measuring degrees of strength of feeling, developed up to the present by the Washington Public Opinion Laboratory. These intensity phrases are scaled out of context and scaled again in diverse contexts, in

[1] *In our standardizing and interdisciplinary S-notation, these scale specifications are denoted by scripts in the corner positions on any variable thus, e.g.,*

$$\begin{array}{rcl} \textit{origin points} = x & & x = \textit{exponents, powers} \\ & X & \\ \textit{units} = x & & x = \textit{indices, homogeneous variables.} \end{array}$$

positive vs. negative contexts, and in seried contexts of the range of degrees along one continuum.

Certain word scales are recommended for use in polls. The user is warned of their current limitations.

Procedures for scaling phrases of degree

The Variables in Scaling

Scaling procedures, in common with all behaviour, may be analyzed here into the three major dimensions of *actors acting* on *objects*—the grammatical human subject, verb, and direct object. The actors in Experiment 1 were forty judges chosen from a university class in scale construction. The objects acted on were the 81 intensity phrases such as " very strongly ", " strongly ", " moderately ", etc. listed in the tables below. This set of phrases specified the universe from which scaled subsets of intensity phrases were sought and developed.

These 81 phrases were in three sets (see Tables 1 and 2) where Set A comprised 23 phrases denoting degrees in general ; Set B comprised 11 phrases of frequency of occurrence ; and Set C comprised 47 phrases in 12 subsets denoting intensity of feeling and related phrases for degrees of " certainty ", " importance ", etc.

Each act was a judging in which each judge placed each phrase, according to its meaning as he perceived it, by a checkmark at some one point on the nine-point graphic and numerical continuum below :

No.	**Phrase to be rated**	**Rating scale for strength**								
						as at				
			less			present			more	
1	" As at present "	1	2	3	4	5	6	7	8	9
2	" None "	1	2	3	4	5	6	7	8	9
3	" Complete "	1	2	3	4	5	6	7	8	9
4	" Slightly less "	1	2	3	4	5	6	7	8	9
5	" Somewhat more ", etc.	1	2	3	4	5	6	7	8	9

The researchers in the research procedure then reacted to these judgment acts as their objects of action by computing the first, second, and third quartile points (Q_1, Q_2, Q_3 respectively) of each phrase on the nine-point rating scale. The second quartile or median point, Q_2, of any phrase was called its scale position, or simply " score ", as used in the Thurstone scaling technique. This was the average strength or intensity denoted by that phrase according to these judges. Thus the phrase, " almost none ", was judged on the average to be at scale position 2.04 ; the phrase, " almost complete ", was judged to be at scale position 8.06 ; while the phrase " as at present ", was judged to be at the exact centre with scale position of 5.00. By rounding off decimals these three phrases may be said to have scores of 2, 8, and 5, respectively. This scale has a range of eight unit intervals from the lowest possible score of 1 to the highest possible score of 9.

The Criteria or Rules for Scaling

From these 81 trios of Q values, subsets of phrases as listed in Table 1 may be chosen for " scalettes ". These scalettes should maximize eight criteria, which have been combined in choosing the sets of phrases recommended for scales below.

The first criterion is to choose unidimensional phrases. This means a set of phrases denoting differing degrees of *one kind.* This is guaranteed in the Cornell technique by a reproducibility above 90 per cent and in the Thurstone technique by the " irrelevance criterion " (Edwards and Kilpatrick, 1948, Guilford, 1936 and Guttman, 1944). Thus the phrases in Set A below refer to amounts in general, while those in Set B refer to a temporal frequency dimension, etc.

The second criterion is to select phrases at *equal* intervals so the scores will increase by even steps. This tends to define operationally a cardinal scale.

The third criterion is to select a set of phrases so as to cover the *whole range* from 1 to 9 as fully as possible.

The fourth criterion is to select phrases each of which has least *ambiguity* or dispersion. The measure of ambiguity used here was the interquartile range, $Q_3 - Q_1$, in which half the judgments must occur. A large interquartile range for a phrase means a large dispersion of the judgments of its position and this reflects its ambiguity to the judges.

A fifth and more semantic criterion is to select phrases of greatest *familiarity* to most people. This increases the generality of the scale. Two indices reflect familiarity. One index is the frequency of the word's use as tabled in a standard word count (Thorndike, 1921). Words among the thousand most-used words in the given language are best. Thus " often " is among the most used 400 words, whereas " without hesitation " is included in only the 10,000 most-used words. A second index of familiarity is brevity which is reflected in the words-per-syllable ratio (Flesch, 1946). This ratio rises as the proportion of short words rises. Short words are chiefly the most familiar Anglo-Saxon monosyllables, " the ", " a ", " and ", " or ", " as ", "if", " no ", etc. If the words-per-syllable ratio is above .7, most of the population will understand ; if it is between .7 and .5, uneducated people begin to drop out ; and if it is much below .5, only the highly educated will understand fully. This words-per-syllable ratio is an excellent index of what percentage of a population is likely to understand a questionnaire—or any prose, for that matter.

A sixth criterion in choosing phrases for a scalette is to choose an origin or reference point which is most familiar as being within the respondent's experience. Often the absolute limits of " none " or " all ", " never " or " always ", will anchor the extremes. Objective reference points like " $10 a day " may be unsuitable because its meaning " much " or " little " varies with the respondent's own income. Then relative adjectives, if scaled, can be more predictive ; a series like " as now, more, much more " may equate individual differences and increase the probability that the respondent will act in life in keeping with his assertion in a poll.

The most familiar origin is suggested by the words " *me here now.*" This takes the respondent himself and his current perception of things in the present tense for the scale origin or chief reference point. It fixes the universally familiar way of thinking for comparisons of degree along the dimensions of time, space, or population. Thus the phrase, " as at present ", is the only phrase which was judged to be at precisely the mid-point ($Q_2 = 5.00$) with minimum possible ambiguity ($Q_3 - Q_1 = .5$).

A seventh criterion is to select phrases for a scalette which are logically symmetric about a middle phrase. Thus " much more, more, as at present, less, much less " form a logically symmetric set, whereas " almost complete, somewhat more, as at present, slightly less, very little " are unsymmetric. The positive and negative sides of the neutral midpoint should have parallel phrasing.

An eighth issue in selecting phrases for a scalette is the number of phrases to use. Should it be a scale of three items or five, or perhaps of seven or nine steps? For precision the larger number of class intervals is better. But in attitude scales many people have difficulty in discriminating beyond five degrees. In oral questioning as in a poll three alternatives are preferable to five. (Note that folk experience, embalmed in language, uses only three degrees in comparing most adjectives and adverbs, namely, the positive, the comparative and the superlative as in " some, more, most.") Experimental work on discriminations in psychology and the recent development, in information theory, of measures of uncertainty or absolute complexity suggest to us that five degrees may be about optimal. Most of our recommended scalettes have just five phrases or scaled anchor points.

If the question is presented to the respondents in print, they can usually cope with more alternatives than their oral spans will permit. Another device to lengthen the oral span is our " tri-bi " or " forkings " technique. By this we ask for choice among three major alternatives and then ask the respondent to subdivide his choice in finer alternatives. Thus the question, " Do you want x as at present, or more of it or less of it? " is followed up, if the answer is " more ", by the question " A little more or a lot more? " The " less " answer is similarly split into two sub-degrees by asking, "A little less or a lot less, would you say? " If a series of " tri-bi " questions occur, the respondent comes to expect them, and the interview moves along quickly with easier deciding from breaking the decisions down into two stages. This technique may be extended as a " tri-tri " forking into three first choices, each subdivided into three second choices. This can get undiscriminating people to discriminate nine degrees where offering them orally nine alternatives simultaneously would completely confuse them. Thus, for example, respondents might first identify themselves as in the upper class, middle class, or lower class, and then further identify themselves as nearer the top, middle, or bottom of that class.

The Context in Scaling

In Experiment 1, Sets A and B, the phrases were judged " out-of-context " while in Set C each phrase was judged in the context of the series of other phrases in its subset. Out-of-context means that the 40 judges were presented with just the phrases in randomized order. In-serial-context means that the phrases of a subset such as " very good, good, fair, bad, very bad " were presented together and in this order.[2]

In Experiment 2 a hundred undergraduate judges judged the scale position in three different content contexts for the phrases presented not randomly but as a series in Subset C 1. An example of the presentation " in context " to the judges is the following :

" If our outlying military bases were attacked, we should go to war."

Answer (check one)	How strongly do you feel about your answer ?	Rank how strongly you feel about your answer on the nine-point continuum from least (1) to most (9). Circle one number.								
		least								most
___Agree	___Very strongly	1	2	3	4	5	6	7	8	9
___Disagree	___Strongly	1	2	3	4	5	6	7	8	9
___Don't know	___Moderately	1	2	3	4	5	6	7	8	9
	___Indifferent	1	2	3	4	5	6	7	8	9
	___Don't know	1	2	3	4	5	6	7	8	9

The resulting scale positions and ambiguities from with-content-and-in-series context judging by these 100 student judges in Experiment 2 are compared in Table 3 and the out-of-context judging by the 40 student judges of Experiment 1 is also summarized in Table 1.

Finally the scale positions and ambiguities of each of these four intensity phrasings on all three issues were computed once for those who agreed with the statement and again for all respondents who disagreed. The question here was whether the intensity ratings would differ as applied to positive vs. negative phrasing of content questions. Does " very strongly " mean a greater deviation from the median position when in the context " I agree very strongly " than when in the context " I disagree very strongly " ?

Findings of scale positions and ambiguities

Out of Context

The results of the above procedure are reported in the tables herewith. Table 1 shows, for the 34 phrases in sets A and B of Experiment 1, the essential data for

[2] *An incompleteness in this experimentation to date is that the phrases of Sets A and B (scaled out of context) were different phrases from those in Set C (scaled in serial context). So it is not possible as yet to compare the out -of-context technique with the in-serial-context technique. Nor are the scale scores of Set C fully comparable with those of Sets A and B. Further research is needed (a) to compare all phrases with each other more rigorously, (b) using larger lists of phrases, (c) on new and larger samples of respondents than in this study.*

applying the criteria[3] for scale construction. The second column gives the recommended score for each phrase on the nine-point scale. It is obtained by rounding the exact medians in column 3 to the nearest integer. This score will tell the poller the degree of strength or amount that is denoted by its phrase on the average. The column medians tell how well the scaler, with whatever number of steps he chooses for his scale, can carry out the second criterion of keeping scale steps at equal intervals combined with the third criterion of covering the range.

The fourth column of Table 1 states the degree of ambiguity of each phrase. The index of ambiguity here is the interquartile range or range of scores which include half the judgments. The minimum possible degree of ambiguity is .5. In our judgment, phrases in these sets with ambiguity less than 1.0 should be preferred and none above 1.5 would be used. These data deal with the fourth criterion for scaling.

The fifth column states indices of simplicity and consequent familiarity for each phrase. These are obtained by dividing the number of words by the number of syllables in each phrase. Phrases with maximal brevity have indices of 1.0. These are monosyllabic words and are the most used words in English (Flesch, 1946). No phrases should be used, we believe, whose simplicity indices are lower than .5, since only a minority of the population will fully understand them. This condemns phrases such as " without hesitation "—which is also commended by extreme ambiguity.

The sixth column states the frequency of use of the most unfamiliar word in each phrase (Thorndike, 1921). This is a measure of its familiarity to the English-speaking public. Words in the first hundred most-used words are best. Words beyond the 2,000-word vocabulary should probably be avoided. The fifth and sixth columns of Table 1 supply the information for the fifth and sixth criteria of scaling discussed above.

The seventh criterion of positive and negative symmetry resulted in tempering the borderline scores of the phrase, " much more ", to make it balance with " much less " and the phrase, " a good deal less ", to make it balance with " a good deal more ".

In-series Context

Table 2 shows the findings for 47 phrases, grouped in series within 12 subsets. These subsets dealt with degrees of intensity or strength of feeling. The scale positions and ambiguities were determined for the phrases presented in a graded series within each subset and not in randomized presentation. The desirable comparison with random order of presentation has not been made. A further weakness

[3] *Table 1 supplies the data for the last seven criteria but not for the first criterion of unidimensionality. Preliminary tests showed that the result of the reproducibility test applied to these word scales was uniformly so well above 0·90 that the full Cornell technique was not applied. The specification that the adjacent phrases in a scale, when presented randomly to the judges, should not have their interquartile ranges overlap, seemed the more rigorous requirement. The phrases will always be used in series which greatly emphasizes their logical unidimensionality as perceived by respondents. Whenever the root phrase is constant and is varied by degree-naming phrases, such as " less," " much less," " very much less," high reproducibility seems inevitable.*

TABLE 1

		Exact score	“ Ambiguity ”		“ Familiarity ”
	Rounded	= “ scale	Interquartile	“ Simplicity ”	= frequency
	score	position ”	range	= words	of use as in
	recom-	= median	= $Q_3 - Q_1$	per syllable	first × words
	mended	= Q_2	(Minimum = 0.5)	(Maximum = 1·0)	× =
SET A *Degree phrases, tested out-of-context*					
complete	9	8.85	0.65	0.50	1,000
almost complete	8	8.06	0.58	0.50	1,000
very much more......	8	8.02	0.61	0.75	100
much more	7	7.67	1.04	1.00	100
a lot more	8	7.50	1.06	1.00	1,000
a good deal more ...	7	7.29	0.98	1.00	1,000
more	6	6.33	1.01	1.00	100
somewhat more	6	6.25	0.93	0.66	2,000
a little more	6	6.00	0.58	0.75	100
slightly more	6	5.99	0.57	0.66	1,500
now	5	5.03	0.53	1.00	100
AS AT PRESENT.........	5	5.00	0.50	0.75	300
slightly less	4	3.97	0.56	0.66	1,500
a little less	4	3.96	0.54	0.75	500
somewhat	4	3.79	0.94	0.66	2,000
less	4	3.64	1.04	1.00	500
much less	3	2.55	1.06	1.00	500
a good deal less	3	2.44	1.11	1.00	1,000
a lot less	2	2.36	1.03	1.00	1,000
very little	2	2.08	0.64	0.50	100
almost none	2	2.04	0.57	0.66	1,000
very much less	2	1.96	0.52	0.75	500
none	1	1.11	0.59	1.00	1,000
SET B *Temporal frequency phrases, tested out-of-context*					
always	9	8.99	0.52	0.5	400
without fail...........	9	8.89	0.61	0.66	1,500
often	7	7.23	1.02	0.5	400
usually	7	7.17	1.36	0.25	1,000
frequently	7	6.92	0.77	0.33	1,500
now and then	5	4.79	1.40	1.00	100
sometimes	5	4.78	1.83	0.5	500
occasionally	4	4.13	2.06	0.20	4,500
seldom	3	2.45	1.05	0.5	2,000
rarely	2	2.08	0.61	0.5	2,500
never	1	1.00	0.50	0.5	300

Scale positions and other data on 34 phrases
(Experiment 1, N = 40)

TABLE 2

	Rounded score recom-mended	Exact score = “ scale position ” = median = Q_2	“ Ambiguity ” Interquartile range = $Q_3 - Q_1$ (Minimum = 0.5)	“ Simplicity ” = words per syllable (Maximum = 1·0)	“ Familiarity ” = frequency of use as in first × words × =
SET C *12 Subsets of intensity phrases, tested in series context*					
Subset C1					
very strongly	9*	8.40	1.04	0.50	300
strongly	7	7.07	0.67	0.50	300
moderately	5	5.24	0.99	0.25	3,500
indifferent	3*	3.70	2.20	0.25	6,618
Subset C2					
very certain............	9	8.55	1.05	0.50	500
certain	7	7.13	1.44	0.50	500
uncertain	3	2.83	2.50	0.33	4,000
not certain	3	2.64	2.62	0.66	500
Subset C3					
extremely vital	9	8.79	0.84	0.40	6,000
very vital..............	8	7.55	1.05	0.50	5,000
vital	6	5.92	1.63	0.50	5,000
insignificant	2	2.12	1.14	0.20	6.047
Subset C4					
of great importance	8	8.05	0.91	0.60	2,500
don't care	5	4.63	2.00	1.00	1,000
doesn't mean anything	2	2.50	2.71	0.50	3,000
Subset C5					
very sure	8	8.15	0.95	0.66	300
sure	6	5.93	1.87	1.00	300
not sure	3	2.82	1.24	1.00	300
Subset C6					
without hesitation ...	8	7.50	6.54	0.33	10,000
with little hesitation	6	5.83	3.40	0.43	10,000
hesitant	5	4.77	1.06	0.33	10,000
with some hesitation	4	4.38	1.60	0.50	10,000
with considerable hesitation	3	3.29	3.39	0.30	10,000
with much hesitation	3	3.20	5.25	0.50	10,000
with great hesitation	2	2.41	6.00	0.50	10,000
only as a last resort	2	1.70	7.30	0.71	3,000

* These scores are tempered slightly outwards from the middle as a compromise between their in-context and out-of-context scale positions.

Scale positions and other data on 47 intensity phrases
(N = 40)

	Rounded score recommended	Exact score = " scale position " = median = Q_2	" Ambiguity " Interquartile range = $Q_3 - Q_1$ (Minimum = 0.5)	" Simplicity " = words per syllable (Maximum = 1·0)	" Familiarity " = frequency of use as in first × words × =
Subset C7					
very good	8	8.08	1.12	0.66	100
good	7	6.72	1.20	1.00	100
fair	5	4.96	0.77	1.00	200
bad	3	2.83	0.93	1.00	500
very bad	2	1.50	1.13	0.66	500
Subset C8					
essential	8	7.58	1.85	0.33	5,000
non-essential	3	2.58	1.67	0.25	5,000
Subset C9					
very urgent	8	8.00	0.90	0.50	5,544
urgent	6	6.41	1.53	0.50	5,544
Subset C10					
very important	8	8.22	1.16	0.40	1,000
important	7	6.83	1.14	0.33	1,000
don't know	5	4.82	0.82	1.00	1,000
undecided	5	4.76	1.06	0.25	10,000
not important	3	3.09	1.33	0.50	1,000
unimportant	3	2.94	1.42	0.25	1,000+
very unimportant	2	1.75	1.25	0.33	1,000+
Subset C11					
object strongly to	4	3.50	6.07	0.60	1,000
feel strongly toward	8	7.80	1.60	0.75	1,000
Subset C12					
very crucial	8	8.29	1.12	0.50	—
crucial	6	6.39	1.73	0.50	—
doesn't make any difference	3	2.83	3.13	0.50	3,000

in the data in Table 2 is that too few phrases were explored. Thus in the Subset C1 there were no alternatives to put in the place of " moderately " and " indifferent " when these turned up as ambiguous, unfamiliar, and not simple. Phrases such as " not strongly " and " not at all strongly " should have been tried also in order to yield a well-graded series that is also unambiguous, simple, and familiar.

TABLE 3

Intensity phrase	Issue	N number of responses	Scale position Q_2	Mean scale position in context	Scale position out of context (from 40 different judges)	Ambiguity $Q_3 - Q_1$
" Very strongly "	1	295	8.96			0.54
	2	197	8.91	8.92	8.40	0.76
	3	161	8.91			0.58
" Strongly "	1	269	7.01			1.28
	2	162	7.20	7.11	7.07	1.22
	3	271	7.12			1.14
" Moderately "	1	305	4.78			1.58
	2	189	4.77	4.82	5.24	1.30
	3	277	4.92			1.40
" Indifferent "		(Insufficient data)				

Stability of intensity phrases in diverse contexts
(Experiment 2, 100 judges and 31 content statements of attitude on three issues)

With Pro-con Content and In-series Context

Table 3 shows the findings from 100 judges for the phrases of Subset C 1 on strength of feeling when presented in a graded series and as applied to 31 scaled statements about the following three issues :

Issue 1 — resistance to starting a war ;
Issue 2 — drafting of women for military service and defence work ;
Issue 3 — amount of government control.

The chief finding here is the close agreement of the scale position and degree of ambiguity of each intensity phrase among the three issues. Apparently this much diversity of context does not appreciably shift the scores of the intensity phrases.

However, these scores disperse a little more in-context (8.92 − 4.82 = 4.10) than out-of-context (8.40 − 5.24 = 3.16).

In Positive or Negative Context

In Table 4 the stability of these intensity phrases of Subset C 1 is shown when in context of negative versus positive content of the basal opinion. Again the stability or agreement is found to be high. The phrases, " very strongly ", " strongly ", etc., have the same meaning as rated whether coupled with the pro-con content, " I agree ", or with " I disagree ".

Again the four intensity phrases show scale positions more dispersed in context than out of context. The in-context range is about 40 per cent greater than their out-of-context range.

TABLE 4

ntensity phrase	Pro or con endorsement	N number of responses	Scale position Q_2	Mean scale position in context	Scale position out of context (from 40 different judges)	Recommended rounded scale score	Ambiguity $Q_3 - Q_1$
ery strongly	" I agree "	299	8.93				0.60
	" I disagree "	514	8.93	8.93	8.40	9	0.58
trongly	" I agree "	430	7.13				1.08
	" I disagree "	469	7.14	7.13	7.07	7	1.22
Moderately	" I agree "	502	4.81				1.64
	" I disagree "	468	4.84	4.83	5.24	5	1.32
ndifferent	" I agree "	32	2.60				1.78
	" I disagree "	42	2.55	2.58	3.70	3	1.46

Stability of intensity phrases in positive and negative contexts
(Experiment 2, 100 judges, 31 content statements of attitude)

The general conclusion seems to be that relative phrases commonly used to show intensity of feeling or other degree of opinion can be scaled. The scores or average scale positions of the phrases did not vary appreciably here from one context to another nor from positive to negative contexts. The median scores dispersed a little more in-context than out-of-context, but these shifts have about the same degree of approximating as the rounding off of the exact median positions to the nearest integer to get the scale score of each phrase. The average adjustment for rounding off scores from either the in-context or out-of-context exact scale positions is less than four per cent of the nine-point range of this scale.

These scale scores are averages in a sample of persons. They are intended for use in polls dealing with samples of a population and will be much less exact for individual prediction. Differences between individuals in judging the phrases are much greater than between samples of individuals or between contexts. In Thurstone scaling the judgments of the scale positions of sentences have been found to transcend the opinions of the judges about the referents of those sentences. Judgments of meanings of words within a language group are less dispersed than the attitudes those words represent. This general experience of attitude testing is expected to mean here that the scale scores of phrases are likely to stay in the order found here when redetermined in other samples of judges or contexts. But this opinion of the authors should be checked by more extensive and rigorous experiments determining word scales for degrees of opinion with increasing exactness.

WORD SCALES RECOMMENDED FOR POLLS

From the findings above the following subsets of phrases may be selected. Each subset is a scale according to the combination of the eight criteria for scaling discussed above. These subsets of phrases of degree seem the best " word scales ", or " scalettes ', that are available from the 81 phrases studied here.

They may be used in polls, with the score on a nine-point range as given below for each phrase, until further studies replace them with better scalettes.[4]

SCALETTE A, *for degrees of strength of feeling :*

Phrase	Score
very strongly	9
strongly	7
moderately	5
indifferent	3

SCALETTE B, *for degrees of sureness :*

very sure	8
sure	6
not sure	3

SCALETTE C, *for degrees of excellence :*

very good	8
good	7
fair	5
bad	3
very bad	2

SCALETTE F, *for degrees in general, 9-point range, 5 equal steps :*

complete	9
much more	7
as at present	5
much less	3
none	1

SCALETTE J 1, *for degrees in general, 7-point range, 5 steps, dense in middle :*

a lot more	8
a little more	6
as at present	5
a little less	4
a lot less	2

SCALETTE D, *for degrees of importance :*

Phrase	Score
very important	8
important	7
unimportant	3
very unimportant	2

SCALETTE E, *for degrees of temporal frequency :*

always	9
often	7
now and then	5
seldom	3
never	1

SCALETTE K, *for degrees in general, 5-point range, 5 equal steps :*

a good deal more	7
a little more	6
as at present	5
a little less	4
a good deal less	3

SCALETTE G, *for degrees in general, 9-point range, 5 steps, dense in middle :*

complete	9
more	6
as now	5
less	4
none	1

[4] *A study has been started comparing the reliability and respondent's ratings of three ways of presenting scales. These ways are the Stapel scalometer (five white or positive boxes above and five black or negative boxes below in a vertical series), a " rank digit " scale (5 digits (or 7 digits in an alternate form) horizontally printed on a card with words anchoring the two ends), and three of the 5-word scalettes. This study will appear in* Journalism Quarterly *in 1960.*

SCALETTE J 2, *variant of J 1 :*

very much more	8
slightly more	6
as at present	5
slightly less	4
very much less	2

SCALETTE J 3, *variant of J 1 :*

very much more	8
somewhat more	6
as at present	5
somewhat less	4
very much less	2

SCALETTE H, *for degrees in general, 9-point range, 7 steps, gaps at 3 and 7 :*

complete	9
very much more	8
a little more	6
as at present	5
a little less	4
very much less	2
none	1

SCALETTE I, *for degrees in general, 9-point range, 9 equal steps :*

complete	9
almost complete	8
much more	7
a little more	6
as at present	5
a little less	4
much less	3
almost none	2
none	1

Variant scalettes may be formed by deleting or adding phrases to the above Scalettes E - K since the scores in these are independent of each other as each was judged by itself out-of-context. Thus if the public's opinions are wholly on one side of the " as at present " point, a shortened and unsymmetric scale may be made up which does not offer the completely unused responses on the other side of that scale.

REFERENCES

CANTRIL, HADLEY (1947). Gauging Public Opinion, Chapter III (Princeton).

DODD, STUART C. (1942). Systematic Social Science (offset edition ; University Bookstore, Seattle).

EDGERTON, H. A. and TOOPS, H. A. (1928). A table for predicting the validity and reliability of a test when lengthened. *J. educ. Research*, **18**, 225.

EDWARDS, A. L. and KILPATRICK, F. P. (1948). A technique for the measurement of social attitudes. *J. appl. Psychol.*, **32**, 374.

FLESCH, R. (1946). The Art of Plain Talk (New York).

GUILFORD, J. P. (1936). Psychometric Methods (New York).

GUTTMAN, L. (1944). A basis for scaling quantitative data. *Amer. Sociol. Rev.*, **9**, 139 .

THORNDIKE, E. L. (1921). The Teacher's Word Book (New York).

SPECTRA OF FRICATIVE NOISE IN HUMAN SPEECH*

PETER STREVENS
University of Edinburgh

This paper describes the results of a spectrographic analysis of a number of voiceless fricatives. The sounds are shown to be capable of description in terms of the frequencies of the lower and upper limits of energy present, the presence or absence of formant-like concentrations of energy, and the over-all relative intensity of the sounds.

The sounds investigated fall into three groups: front, mid and back, corresponding to the regions of the vocal tract within which they are produced. Sounds in the front group have a long spectrum, with little patterning of peaks of energy; their relative intensity is low. Sounds in the mid group have a short spectrum, with the main region of energy at a higher frequency than in the other groups; their relative intensity is high. Sounds in the back group have a spectrum of medium length, exhibiting a formant-like patterning of energy; their relative intensity is intermediate between the other groups. Tentative criteria are advanced for distinguishing between members of each group.

Combining this evidence with general phonetic knowledge it is possible to make a number of statements about other categories of sounds which include a component of fricative noise: i.e. voiced fricatives, stops, and affricates.

PRELIMINARY DISCUSSION

The phonetic category of voiceless fricatives comprises speech sounds consisting solely of turbulent noise, or *hiss*.

Nine different voiceless fricatives were selected for analysis in the present study. They are only some of the total members of the class, and were selected to provide a wide coverage of different places of articulation and shapes of orifice. The sounds are those commonly referred to by the following symbols and articulatory labels:—
Φ (bi-labial; f (labio-dental); θ (dental); s (alveolar); ʃ (palato-alveolar); ç (palatal); x (velar); χ (uvular); h (glottal).
It should be made clear that the term *voiceless* is used throughout this paper in its normal, general phonetic sense. Voiceless fricatives are sounds produced with no vibration of the vocal cords. Their spectrum is basically that of aperiodic random noise.

The author has been concerned in the operation for research purposes of PAT, the parametric artificial talking device designed by W. Lawrence. (Lawrence, 1953; Strevens, 1958a, 1958b; Strevens and Anthony, 1958). It was clear from an early stage that the work must include the adequate simulation of at least the voiceless fricatives of English. In practice it was found that the available data were not sufficient

* *This work was undertaken under a contract between the Ministry of Supply and the University of Edinburgh for research into the specification of speech by means of acoustic parameters.*

to programme PAT to do this. The investigation now described was undertaken with a view to providing the basic data for the purpose.

Voiceless fricatives are all produced by turbulent air-flow caused by a constriction in the vocal tract at some point in or above the larynx. The constriction may vary as to position in the tract, degree of constriction, area of constriction, and shape of orifice. Further, although it is customary to think of the vocal tract as if it were a tube having a cylindrical cross-section, it must not be forgotten that the tract is in places highly mobile, and that it may alter its shape to a considerable extent independently of constrictions such as those under discussion. Finally, the air-stream may vary as to pressure or rate of flow. A variation of any combination of these factors can be expected to cause a variation in the physical nature of the resulting sound in any of three ways: by altering the spectrum of the original source of sound ; by altering the filter function of the tract as a whole ; by altering the intensity of the acoustic energy produced.

The spectrum of the sound-source will depend on the degree and the area of constriction, on the shape of the orifice or orifices, and to a minor extent on the rate of air-flow. The chief effect of increased air-flow is to increase the overall acoustic energy. The filter function will depend largely (though not wholly) on the position of the constriction within the vocal tract, since this position will decide what portions of the vocal tract are contributing to the shaping of the source spectrum.

Because of the inter-dependence of the physiological and the physical events it is necessary to consider them both ; the following order will be employed: first, the modifications of the vocal tract which impede the air-stream and give rise to turbulent flow ; secondly, variations in air-pressure and their relation to acoustic intensity ; thirdly, the spectrographic analysis of voiceless fricatives ; fourthly, an extension of the results thus obtained to consideration of other sounds containing a component of noise.

Modifications of the vocal tract

The position of the constriction within the vocal tract is possibly the chief cause of identifiable differences of sound quality in voiceless fricatives. (This is indeed true of all consonants ; a traditional method of defining consonants is in terms of " place of articulation ".) The organs of articulation concerned in the production of the nine selected voiceless fricatives /Φ f θ s ʃ ç x χ h/ and the typical positions within the vocal tract at which the constriction (and hence the turbulent air-flow) occurs, are as follows:

/Φ/ is produced with the constriction at the lips. Its occurrence in British forms of English is limited to certain Scottish and Irish dialects and to the exclamation " Phew ! "

/f/ is produced with the upper teeth close to the inner surface of the lower lip. The air-stream passes between the teeth and the lower lip, and also through some of the interstices between the upper teeth.

/θ/ is produced with the tip of the tongue close to, or touching, the inner edge of the upper incisors. However, there is a good deal of personal variation in the place of articulation of this sound: it has been described as occurring with the tip of the tongue protruding between the upper and lower teeth ; but a tip-teeth articulation is believed to be the commonest one.

/s/ is produced with either the tip or the blade of the tongue raised to approach the alveolar ridge.

/ʃ/ is produced with the blade of the tongue, or the tip and blade, approaching the palate approximately at the part where the alveolar ridge merges into the main body of the hard palate.

/ç/ is produced with the front of the tongue (i.e. roughly that part of its surface which opposes the dome of the palate when the tongue is at rest) approaching the hard palate somewhat forward of its highest point.

/x/ is produced with the back of the tongue approaching the midle of the soft palate.

/χ/ is produced with the back of the tongue approaching the back of the soft palate and the back wall of the pharynx.

/h/ is the subject of some controversy. It is thought by many that the turbulent air-flow is produced somewhere in the larynx ; others believe " cavity-friction " to be generated throughout the vocal tract. The exact mechanism is not clearly understood.

Each of the articulations thus briefly described occurs at a different point in the vocal tract and can therefore be expected to lead to a different shaping of the source spectrum. The source spectrum itself will differ to some extent according to the nature of the orifice formed at the place of constriction. A short description of typical orifices for each of the nine sounds will demonstrate the occurrence of a wide variety of orifice shapes.

/Φ/ is produced with a long narrow slit between the lips.

/f/ is produced with a narrow opening between the upper teeth and the lower lips. There may also be a contribution through slits between the teeth.

/θ/ is produced with a narrow slit between the bottom of the upper teeth and the surface of the tongue ; the configuration of the teeth affects the quality of the sound produced.

/s/ is produced with a narrow slit which may sometimes be accompanied by a deep groove and pit in the tongue.

/ʃ/ is produced with a wider slit or groove than for /s/ (with a greater area of turbulence). Further, the main body of the tongue assumes a different posture for /ʃ/ than for /s/.

In the production of /ç x χ/ the constriction is far back in the mouth. The chief difference between them is the place at which the constriction occurs.

/h/ is produced with increased air-flow through the larynx ; the area of turbulence is probably very extensive.

The above set of statements is to be taken as an approximation and a normalisation, and is not presented as an absolute or final description. Palatographic studies, especially those using the direct photography method (Anthony, 1954 ; Abercrombie, 1956 ; Ladefoged, 1957 ; Way, 1957) show that there are often large variations of the shape of the orifice and even of the place of constriction, within the speech of a single individual as well as between different speakers. Further, the description of these items by no means exhausts even commonly-observed methods of production of hiss. The air-flow in most cases passes over the median line of the tongue, but it may for some voiceless fricatives pass over one or both sides of the tongue ; the tongue may be grooved or flat, or " pitted " ; there may be more than one constriction at one time ; the nasal cavity may be coupled-in with an air-flow sufficient to cause nasal turbulence, and so on.

It is clear that there are available to the speaker compensatory processes which enable him to produce an acceptable quality of voiceless fricative using quite a variety of different articulatory postures. (These processes are familiar to all those who have lost or acquired teeth.) The foregoing descriptions, then, are a catalogue of some configurations that *do* occur, not of those that *must* be used.

VARIATIONS OF AIR-FLOW

A factor which must now be studied is the acoustic *intensity* of these sounds.

Two assumptions underlie this section of the paper ; first, that variations in the air-flow of speech have a major effect upon the *intensity* but only a negligible effect upon the *spectrum* of the sound produced ; secondly, that the amount of acoustic energy produced during a voiceless fricative is closely related to the rate of air-flow involved.

It is common observation that, to put it roughly, all the voiceless fricatives occur sometimes loud and sometimes soft, but some are loud most of the time while others are soft most of the time. It is of importance to discover whether this is because of inherent differences of acoustic energy or because of other factors such as those arising in the initiation and modification of the pulmonic air-streams.

In English (and presumably in all languages using only an egressive pulmonic air-stream) the pressure of the expiratory air is operated upon by two variables: (1) the *affective variations* of pressure ; (2) variations of *phonetic impedance*. These labels need brief explanation:—

(1) During most speech, the mean pulmonic air pressure remains relatively constant. Fig. 1 shows the relation between the sub-glottal pressure and the volume of air in the lungs during a normal conversational utterance. As speech begins the air-pressure rises to an appropriate level (about 3 cms. aq. for normal speech). The pressure-level appropriate to shouting is higher than that for quiet speech. Pressure levels above the minimum for speech are decided by the general degree of loudness at which the speech is to be produced. The monitoring of this factor is entirely subjective and automatic ; the term " affective variations " is proposed, to describe the gross changes in air-pressure level which produce the required loudness.

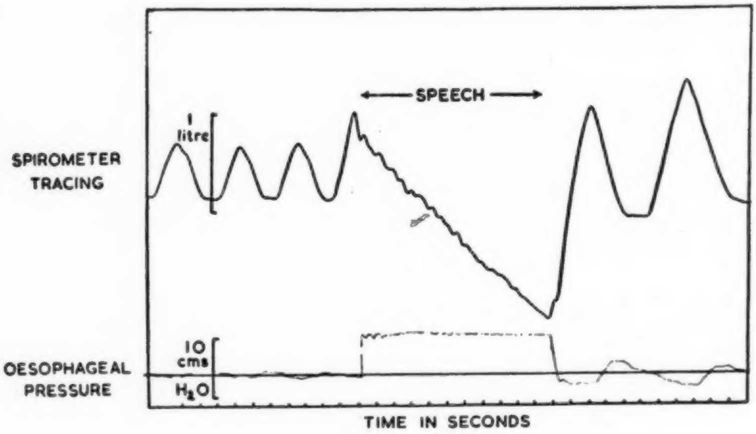


Fig. 1. Two parameters of the air-stream used in respiration and speech. Simultaneous records of volume and pressure. (By courtesy of P. Ladefoged.)

(2) The load on the air-stream varies in complex ways. The air-stream itself above the larynx has two modes of flow ; breath and voice, the former being normally at a higher total rate of flow than the latter. In either mode the flow may be either free or impeded to some extent. Both the extent and the duration of the obstruction will vary from one sound to another. Stops have a momentary but complete obstruction—an infinite phonetic impedance—while fricatives have a lower phonetic impedance of a longer duration. These changes in impedance are added to the changes in pressure brought about by muscular activity in the form of " chest pulses " (Draper, Ladefoged and Whitteridge, 1957, 1959 ; Ladefoged, Draper and Whitteridge, 1958). Between them they produce rapidly fluctuating peaks of pressure super-imposed upon the speech level.

It is now appropriate to describe some preliminary attempts to relate acoustic energy to tracheal air-pressure.

A microphone, set at a distance of 1 ft. from the speaker's mouth, was connected through a pre-amplifier to a valve millivoltmeter. Variations in the acoustic intensity of the nine voiceless fricatives could thus be read as variations in voltage. A nasal catheter tube terminating in a small, lightly-inflated rubber bulb, was passed into the oesophagus of the two subjects, Peter Ladefoged and the author.

Experiments have shown (Draper, Ladefoged and Whitteridge, 1959) that since the oesophagus is separated from the trachea by only a thin membrane, the variations of pressure on the air-filled bulb could be treated as variations in the pressure of the pulmonic air-stream. Further, the oesophagus is on the lung-ward side of all the places of constriction of the nine sounds under investigation. The catheter tube was connected to a transducer incorporating a movable anode whose output was taken to a second voltmeter. The apparatus as a whole can be calibrated by a manometric device.

The subjects spoke a large number of examples of each of the nine voiceless fricatives, using degrees of muscular effort ranging from shouting to little more than a whisper. The two voltages corresponding to acoustic intensity and air-pressure were read simultaneously for each test item. It was found that peristalsis of the alimentary tract interfered increasingly with experimental readings as the next meal-time approached. Nevertheless sufficient readings were taken for the following points to become clear: (a) similar articulations could be made to produce acoustically different sounds; (b) after some little practice, a subject could learn to vary the intensity and pressure readings to some degree independently of each other; (c) a given sound may be produced habitually by a given speaker at a much higher or lower level of intensity than is used by another speaker.

The intensity reading of each item was divided by the pressure reading of each item. In this way it was possible to arrive at a single index for each sound, and a mean for all examples of a given sound by a given subject. There were strong correspondences between the figures of the two subjects; the indications are that the nine sounds as produced by these two subjects in these tests may be arranged as in Table 1:

TABLE 1

1	(lowest) Φ
2	θ
3	f
4	χ
5	s
6	x
7	ʃ
8	h
9	(highest) ç

Rank order of intensity per unit air-pressure.

The results of this preliminary experiment should be regarded with reserve, until there has been an opportunity to verify them using a more rigid experimental procedure and a larger number of subjects. When this has been done, it will be possible to establish an order (which may be similar to that given above) representing the relative intensities of the different fricatives when they occur in a sample of speech with a given mean pulmonic air-pressure level, e.g., as shown in Fig. 1.

THE PROGRAMME OF SPECTROGRAPHIC ANALYSIS

Spectrographic analysis had not been attempted in the few papers previously published on these sounds. (Halle, Hughes and Radley, 1957 ; Hughes and Halle, 1956 ; Meyer-Eppler, 1956.) A small pilot investigation was undertaken with a view to discovering what the drawbacks would be in using spectrographic analysis, whether adequate data could be obtained, and what procedures should be used.

It was found at once that voiceless fricatives occurring in connected speech rarely gave usable spectral information, for two reasons: first, the over-all acoustic energy of the voiceless fricatives is generally much lower than for the stressed vowels by whose peaks the signal level is usually adjusted, consequently the full spectral pattern of the fricatives is too weak to appear ; secondly the duration of these items is often quite short, so that the quantity of pattern available for study is inadequate. The apparatus available for this study included a Kay Sonagraph ; the first practical task was to find a technique of using the instrument which would overcome these difficulties. A sample of speech was recorded with the aim of studying the voiceless fricatives which occurred in that utterance ; the expedient was tried of setting the recording level by these fricative items, and not by the peaks of energy occurring during vowels. This meant that the stressed vowels and may other voiced sounds were badly overloaded, but it was immediately apparent that greater detail was visible during the voiceless fricatives, and that the patterns for a given fricative were consistent between one utterance and another.

To overcome the problem of the short duration of naturally-occurring fricatives the possibility was considered of using the sounds in isolation and of deliberately lengthening them. Given good listening conditions (e.g., the close proximity of speakers and the low background noise usual for quiet conversation) listeners experienced no difficulty in identifying voiceless fricatives spoken in isolation even when no visual clues were present. Lengthening the sounds in isolation only made their identification more immediate and certain. A scheme of investigation was therefore prepared on the assumption that the spectra of isolated, lengthened utterances of the voiceless fricatives would contain all the clues necessary for their auditory identification, and would not contain any undue quantity of spurious components. (The voiceless fricatives occur more frequently with a short duration, but the additional length used in this investigation is not necessarily an unreal factor. The following occurrences of lengthened voiceless fricatives are relatively common: /ΦΦΦ/ when expressing relief after a narrow escape ; /sss/ expressing disapproval at the theatre ; /fff/ when talking to children (" Pufff, puffff . . . . " etc.) ; /ʃʃʃ/ asking for silence, etc.)

The list of items was selected, as already mentioned, so as to cover a reasonable range of possible places of production within the vocal tract. Thirteen past and present staff and post-graduate students of the Phonetics Department of Edinburgh University acted as subjects. Each subject spoke each of the items in isolation, lengthened to approximately one second. Using as subjects people with a professional training in phonetics enabled satisfactory recordings to be obtained very quickly with the minimum of

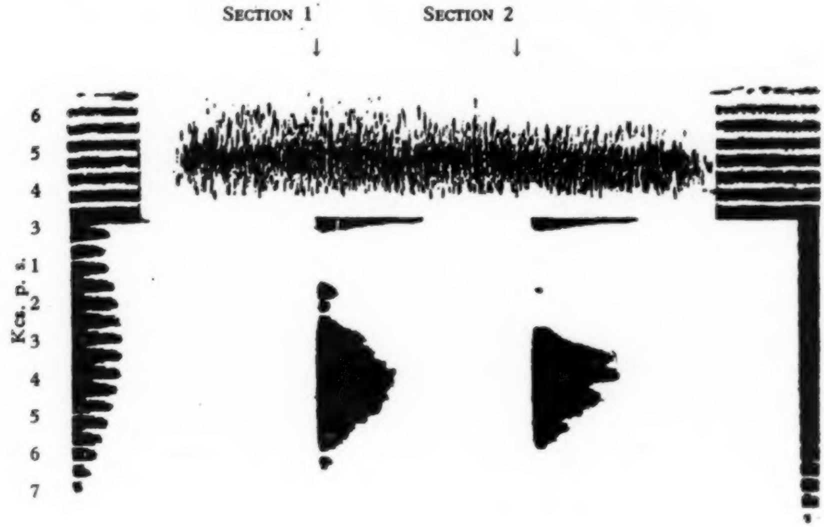


Fig. 2. A typical pair of sections: the /s/ of speaker E.

instructions and rehearsal. The utterances were checked both at the time they were made and from the recordings to ensure that the instructions had been carried out. The recordings were made on a Ferrograph tape recorder, with a flat frequency response from 50 to 10,000 cps. ± 2 db, running at 15 inches per second. The recording technique was critical: the subject had to be close enough to the microphone to provide an adequate signal, yet his breath-stream had to be directed in such a direction as not to impinge on the microphone.

Once the recordings were complete the spectrographic analysis was begun. Various displays were tried. The best results compatible with the time and labour involved were obtained from two spectrograms per utterance ; (i) a broad-band spectrogram of the normal scale of 0-8000 cps.; (ii) two amplitude sections of each utterance, chosen one from the first half and the other from the second half. A typical pair of spectrograms is shown in Fig. 2.

Each recording was followed by a calibration tone consisting of a short ' pip ' of square waves from a signal generator set at a frequency of 500 cps. The display given by this ' pip ' has been arranged to provide a calibration scale at each edge of the spectrogram.

Spectrograms of each of these three types were made for each of the nine utterances of each of the thirteen speakers. The spectrograms were then inspected visually. Two things were immediately apparent: first, there were systematic variations of pattern between one item and another in the analysis of a given speaker ; secondly, there were similarities of pattern within a given item in the analyses of all speakers.

Many alternative methods were tried in a search for a simple presentation of the patterns which occur in these spectrograms. The least unsatisfactory is to produce an " average line spectrum " for each utterance. These line spectra indicate the range of frequencies within which energy is shown to be present on the spectrograms. The variation of upper and lower limits of frequency within any one spectrogram is surprisingly small, and once it was decided that extremes of variation were to be ignored and only typical frequency limits shown, the preparation of line spectra was easy.

On these lines cross-bars are marked at frequencies when peaks of energy occur. No distinction is made between peaks of different height, or of different breadth: the cross-bars mark simply frequencies at which peaks of some kind occur. The reason for not indicating the magnitude of the peaks is a purely practical one: there is no simple way of describing them. It is arbitrary enough to decide by visual inspection that a peak is or is not present, but distinctions between different sizes or shapes of peak are not feasible, as anyone will confirm who has studied fricative spectra.

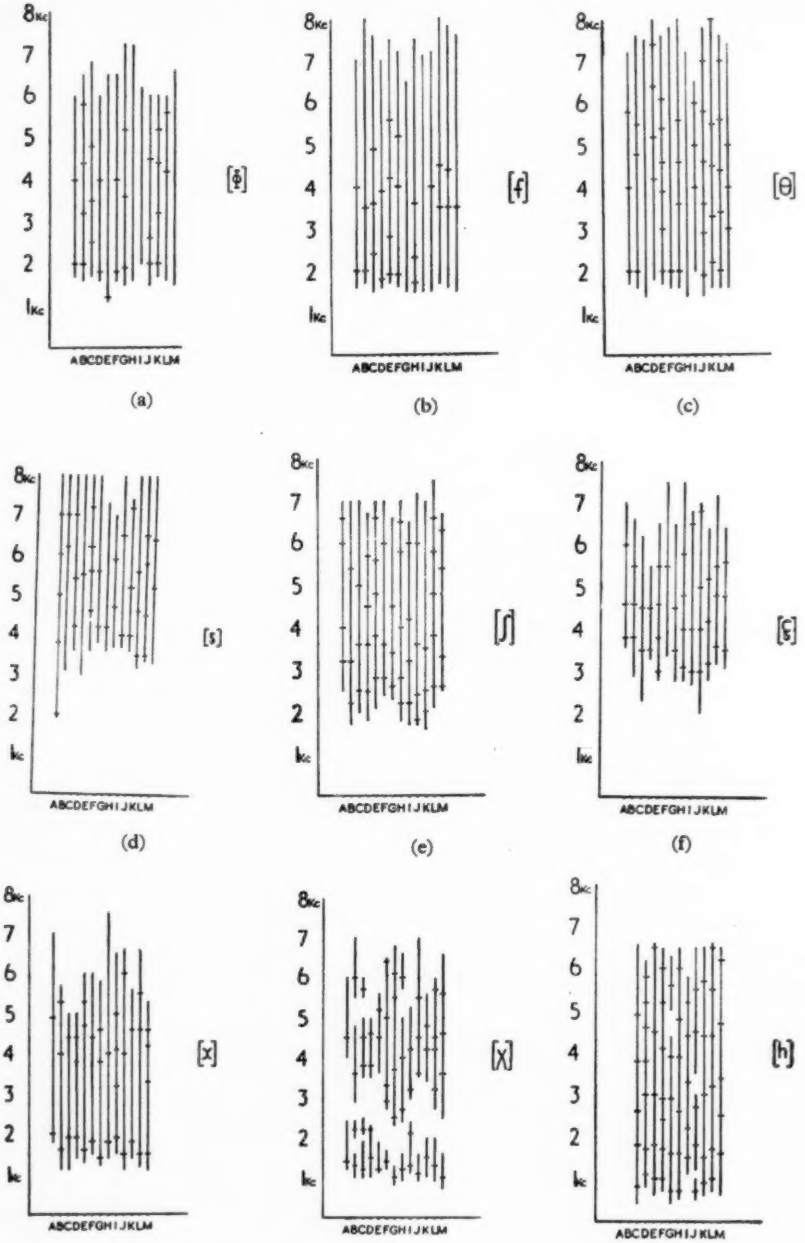
Here it may be mentioned that the amplitude cross-sections were considerably less helpful than might have been imagined. One single cross-section per fricative yields apparently helpful information about peaks of energy. A second cross-section from the same utterances, however, almost invariably gives information that conflicts in its details with the previous analysis. The operative point is that the conflict concerns the *details*: the rough outline is generally similar. Presumably the reason for the discrepancies is that we are dealing with the acoustic shaping of aperiodic noise. The general aspect of the cross-section is determined by the shaping which in turn is conferred by the configuration of the vocal tract ; but the random nature of the sound is illustrated by variations of detail from one instant to another. To sum up, one cross-section per utterance gave a spurious appearance of firm detail, while two per utterance gave conflicting evidence. The line spectra and the cross-bars indicating frequency were therefore compiled chiefly from the broad-band spectrograms.

These line spectra are shown as Fig. 3, diagrams a to i.

No account can be taken of the upper limits of frequency above 8 Kcs., since the frequency response of the spectrograph dropped away sharply above that point. From rough tests made with recordings played at half speed and then given the same analysis it appears that in the spectra shown as reaching 8 Kcs., some energy is in fact present in most cases up to 10 Kcs., and in a few cases up to at least 12 Kcs. But the evidence is not sufficient to be presented here and no systematic study of the upper limits of frequency has been attempted, where these lie above 8 Kcs.

The averaged line spectra shown in the diagrams must be accompanied by a verbal description of the average spectra of each of the nine voiceless fricatives ; this will include a statement of intensity. By this is meant the order of ranking from the investigation of intensity per unit pressure described above. It must be repeated that no general validity is claimed for this figure.

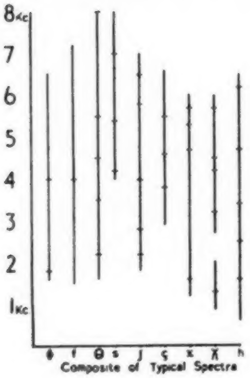
1. /Φ/ (Fig. 3a) Lowest frequency at which energy is visible on the spectrogram is between 1600 and 1650 cps. Low peaks of energy tend to occur around 1800 - 2000 cps., 4000 - 4500 cps., and 5500 cps. Energy rarely above 6500 cps. Intensity ranking: lowest of 9.
2. /f/ (Fig. 3b) Lowest frequency is around 1500 - 1700 cps. Low peaks of energy tend to occur around 1900 cps., 4000 cps., and occasionally 5000 cps. Upper limit of frequency is rarely below 7000 cps., usually around 7500 cps. In general, a higher upper limit than No. 1. Intensity ranking: 3rd in ascending order.
3. /θ/ (Fig. 3c) Lowest frequency varies, but lies between 1400 and 2000 cps. Low peaks of energy tend to occur, the lowest being close to 2000 cps., the upper peaks varying somewhat, but tending to lie about 1000 cycles apart. Upper limit of frequency rarely below 7200 cps.; some speakers reach 8000 cps. In general, a somewhat higher upper limit than No. 2. Intensity ranking: 2nd in ascending order.
4. /s/ (Fig. 3d) Lowest frequency almost always above 3500 cps. Peaks of energy tend to occur with no apparent pattern, except that they do not lie closer to one another than 1000 cycles. Upper limit of frequency exceeds 8000 cps. in most cases. Intensity ranking: 5th in ascending order.
5. /ʃ/ (Fig. 3e) Lowest frequency varies between 1600 and 2500 cps. Peaks of energy tend to occur not less than 1000 cycles apart and the aspect of amplitude cross-sections shows a weighting towards the bottom of the pattern. Upper limit of frequency shows a sharp cut-off around 7000 cps. Intensity ranking: 7th in ascending order.
6. /ç/ (Fig. 3f) Lower limit of frequency varies generally between 2800 and 3600 cps. Peaks of energy tend to appear at roughly 1000 cycle intervals ; these peaks are sharper than those in No. 5. Upper frequency limit very variable, but usually between 6000 and 7200 cps., i.e. lower than for either No. 4 or No. 5. The general shape of the spectrum is like that of No. 4 /s/, but with all values transposed 1000 cps. down. Intensity ranking: greatest of all 9 items.
7. /x/ (Fig. 3g) Lower limit of frequency usually between 1200 cps. and 1500 cps. There is always a strong peak of energy below 2000 cps., with others above about 3500 cps. The aspect of amplitude cross-sections gives a hint of formant-like structure ; the low peak is steeper than the upper peaks, which are often double peaks, some 500 - 600 cycles apart. Upper frequency limit is very variable, usually between 5000 cps. and 7500 cps. A considerable variety of different sound qualities was obtained from the subjects, as was to be expected. Versions judged in phonetic terms to have a more back place of articulation tended to approach more closely to a formant-like structure. Intensity rating: 6th in ascending order.
8. /χ/ (Fig. 3h) Lower limit of frequency varies between 700 cps. and 1200 cps. All spectra bear a marked resemblance to vowels, with a " formant " of one or two



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)



(j)

Fig. 3. Typical line spectra for nine voiceless fricatives. Diagrams (a)-(i) show average line spectra for each item of each speaker ; diagram (j) is composed of one typical spectrum of each item, for comparison.

high peaks between 1000 cps. and 2400 cps. Sometimes there are 3 or even 4 " formants " altogether, with 1 or 2 of them having rather high peaks of intensity, in the region from 3000 cps. to 6000 cps. At first glance the spectrographic pattern is that of a vowel rather than a voiceless fricative. Upper limit of frequency variable between 6000 cps. and 7000 cps. Intensity rating: 4th in ascending order.

9. /h/ (Fig. 3i) Lower limit of frequency usually varies between 400 cps. and 700 cps. The peaks of intensity which occur are so marked as to suggest a multi-formant vowel. One major peak occurs around 1000 cps., one around 1700 cps. At least 5 major peaks occur in each pattern ; spectra for women subjects exhibit more of these peaks than for men. Upper limit of frequency is usually around 6500 cps. Intensity ranking: 8th in ascending order, but for technical reasons the data (and thus the ranking) for this item are suspect.

A consideration of the data provides a basis for distinguishing three groups of voiceless fricatives, the members of each group sharing certain features. It happens (not surprisingly) that the groupings reflect major differences in place of articulation. The groups may therefore be labelled as follows: *Front* (containing the labial and dental sounds /Φ f θ/) ; *Mid* (containing the alveolar and palatal sounds /s ʃ ç/) and *Back* (containing the velar, uvular and pharyngal sounds /x χ h/). The groups are distinguished in terms of spectral pattern in the following way:—

Front Group /Φ f θ/ Long spectrum, covering a range of some 5000 to 6000 cycles ; a peak of energy frequently occurs below 2000 cps., but in general the peaks are un-patterned and " spiky " ; the relative intensity is the lowest of the three groups.

Mid Group /s ʃ ç/ Short spectrum, covering a range of some 3000 to 4000 cycles ; one or more major " humps " around the middle of the pattern ; the relative intensity is the highest of the three groups.

Back Group /x χ h/ Medium spectrum covering a range of some 4000 to 5500 cycles ; a marked " formant-like " structure with invariably a major peak or " formant " around 1500 cps.; the relative intensity is the middle of the three groups.

The above characteristics identify the groups and can be stated with some confidence. The factors which distinguish members within a group are not so clear-cut, and are put forward with some reserve. It is suggested that it is within these groups, rather than between them, that confusions and misidentifications most frequently occur under conditions of restricted efficiency of communication (e.g., on poor telephone circuits). The distinctions between different members of a group may reside in the following criteria:—

Front Group : " Centre of gravity ". It seems from a study of the amplitude cross-sections that the sequence /Φ f θ/ (that is, going progressively further back in place of articulation) displays an increasing weighting of the upper end of the spectrum, accompanied to some extent by a higher upper limit of frequency.

Mid Group : A combination of upper and lower frequency limits. The sequence /s ʃ ç/ (that is, going progressively further back in place of articulation) is accompanied by a change of upper limit from highest to lowest ; the lower limit changes in the same sequence from highest to lowest to intermediate.

Back Group : Doubtful: possibly the frequency of lower limit of frequency, which becomes progressively lower in the sequence /x χ h/ (that is, progressively further back in place of articulation). The foregoing features are summarised in Table II.

The sounds with which this investigation has been concerned are well known as being aperiodic in nature. Unfortunately it has often been assumed that they should be equated with the classical case of random, aperiodic vibration, known as " white noise ". It is now quite clear that the randomness of the sound source is greatly modified by the acoustic shaping characteristics of the vocal tract, and that this shaping varies in several ways. To extend the " white noise " metaphor it might be said that voiceless fricative speech sounds consist of " grey noise, streaked with white and black ".

Extrapolations

Once the analysis and classification of voiceless fricatives has been performed on lengthened, isolated segments, it becomes possible to recognise many or all the features already described from their occurrence in spectrograms of normal connected speech. More important, the same sets of features begin to be seen, at least partially, in spectrograms of voiced fricatives, voiceless and voiced stops, and voiceless and voiced fricatives.

No detailed analysis of any of these sounds has been attempted by the author but

TABLE 2

SOUNDS	ARTICULATION GROUP	RELATIVE INTENSITY	SPECTRUM LENGTH	DISTINCTION BETWEEN MEMBERS OF THE GROUP
Φ	FRONT (Labial & dental)	LOW	LONG (5000-6000 cycles)	Φ lowest " centre of gravity "
f				f intermediate " centre of gravity "
θ				θ highest " centre of gravity "
s	MID (pre-velar)	HIGH	SHORT (3000-4000 cycles)	s highest bottom limit, highest top limit of frequency
ʃ				ʃ lowest bottom limit, intermediate top limit of frequency
ç				ç intermediate bottom limit, lowest top limit of frequency
x	BACK	MEDIUM	MEDIUM (4000-5500 cycles) with " formant-like " structure	x highest bottom limit of frequency
χ				χ intermediate bottom limit of frequency
h				h lowest bottom limit of frequency

The special characteristics of nine voiceless fricatives.

a combination of general phonetic considerations with extrapolations from the data on voiceless fricatives would lead one to expect with considerable confidence the existence of certain predictable features. The following paragraphs summarise these expectations.

We have seen that when turbulent air-flow occurs its spectrum will be related to the place where it occurs, the shape of the orifice concerned and the flow of air through the constriction. If other sounds containing a component of fricative noise are now considered in the light of the data described above, strong indications may be seen as to the probable nature of the hiss. It is convenient to discuss these sounds in categories according to their method of production.

Voiced fricatives (e.g.: /β v ð z ʒ j ɣ ʁ ɦ/)

These sounds are made up of two components: a component of hiss and a component of " vocal tone " or *voice.* It can reasonably be assumed that the acoustic characteristics of the hiss will correspond in most respects to those of the voiceless fricatives. The major difference in articulation is that in voiced fricatives for a given air-pressure the *air-flow* is less than for the voiceless items, since the breath stream is being interrupted and reduced in flow by the action of the vocal cords. For a given air-pressure the acoustic intensity of the hiss component of voiced fricatives is inherently less than that of the corresponding voiceless items.

Evidence tending to confirm this may be found in spectrograms of ordinary speech. It is instructive to compare two amplitude displays (not cross-sections) of the same utterance, using a flat amplifier response for one and high frequency emphasis for the other. The amplitude of the trace during voiceless fricatives is greatly increased by the high frequency emphasis ; the voiced fricatives, on the other hand, are only slightly higher than in the flat-response condition. The energy present at the higher frequencies is clearly less in the voiced fricatives than in the voiceless ones.

Nevertheless hiss remains comparatively easy to identify, even at low intensities and when accompanied by vocal cord vibration. In teaching general phonetics, there is little difficulty in teaching students to identify the presence of hiss, and the hiss seems not to be masked by voice.

Stops

(i) *Voiceless*

Many languages contain voiceless stops, in the production of which the air-stream is momentarily obstructed, then released with a " burst " of fricative noise, more or less short in duration. The spectrum of the hiss on any given occasion will inevitably be like the spectrum of a closely homorganic fricative. Thus in English when the voiceless stops (/p t k/ are released with a burst of hiss, the spectrum of this affrication must be virtually identical with /Φ s x/ respectively, since /Φ/ and /p/ are homorganic, as are /s/ and /t/ and also /x/ and /k/). This is what one tends to find on close examination of suitable spectrograms. (See also Fischer-Jorgensen, 1954.)

(ii) *Voiced*

Perceptible affrication after voiced stops is much less common, at least in English. The reduced rate of air-flow resulting from vibration of the vocal cords causes a smaller build-up of pressure behind the occlusion, duration for duration, than in the case of voiceless stops. This means that when the stop is released the air-flow may be insufficient to cause audible friction. Even when hiss does occur, it is both short in duration and extremely low in level compared with the voiced components. The spectrum is identical with that of the voiceless counterpart.

Affricates

(i) *Voiceless.* Affricates combine in sequence some of the articulatory features of the stops (e.g., complete but momentary obstruction of the air-stream) with other features characteristic of the fricatives (e.g., partial obstruction of the air-stream). An important additional point is that the place of articulation of the stop release is frequently not the same as the place of the fricative articulation within the same sound. The relative levels of the stop release and fricative portions may also be different. Consequently the hiss portion of an affricate may consist of two segments having different spectra. Thus /tʃ/ in *church* begins with a short " stop-release " burst having a spectrum closely similar to /s/ and is followed without a break by a longer " fricative " segment having the spectrum of /ʃ/.

(ii) *Voiced.* In voiced affricates (as in *judge*) the considerations for voiced stops and voiced fricatives apply. The stop-release portion will be at a much lower level of intensity, both absolutely and relatively, than in a voiceless plosive. The " fricative " portion will have audible friction with a spectrum appropriate to its place of articulation at a relatively lower level of intensity.

Fricative sounds in speech synthesis

Although the foregoing remarks on hiss in sounds other than voiceless fricatives have been presented as a sequel to the discussion of fricatives they also received some practical attention. In the preparation of synthetic speech utterances for Lawrence's PAT, the following observations have been made:

(i) variations of hiss spectrum, in the general direction of simulating the spectra described above, lead to improved acceptability and intelligibility of synthetic speech ;

(ii) even if the transitions have been faithfully simulated, voiced fricatives provided with some hiss of the appropriate spectrum are much more acceptable than the same sounds without hiss or with hiss of some different spectrum ;

(iii) in synthesising affricates a great improvement is obtained by working to a pattern of the appropriate " stop-release " spectrum portion, plus the appropriate fricative spectrum. These *ad hoc* observations are no substitute for a careful series of experiments in the synthesis of hiss sounds, but they provide much empirical confirmation and no refutation of the general validity of the foregoing remarks.

Furthermore, the quality of synthetic speech which PAT can produce is already very greatly improved as a result of this analysis of the acoustic spectra of voiceless fricatives; it is now at last possible to write some physical specification of the sounds that the machine is to be instructed to simulate.

Experiments on the identification of stops and of fricatives have been described previously by several writers, notably by workers at the Haskins Laboratories (Liberman, Delattre and Cooper, 1952 ; Cooper, Delattre, Liberman and Gerstman, 1952 ; Liberman, Delattre, Cooper and Gerstman, 1957). The fricative noise used in the experiments described consists of white noise of 600 cycles spectrum width ; the only systematic variations reported are variations of the centre frequency of the band of short-spectrum hiss. (The exception is Dr. Harris (1956, 1958) who used wide-band white noise in some experiments and recordings of human fricatives in others, and experimented with variations in the relative intensity of vowel and fricative.)

The question arises whether substantially different results might be expected from these synthetic speech experiments if they were to be repeated using hiss spectra more closely resembling those found in human speech. The conclusions of Schatz (1954) support the contention that the fricative portions of voiceless stops are in many circumstances sufficient clues for identification, without reference to vowel transitions. It seems probable, therefore, that listeners' judgements on any occasion will be influenced in the direction of the naturally-occurring hiss spectrum most closely resembling the synthetic hiss.

Acknowledgements

The foregoing spectrographic investigations and the descriptions of articulatory and physiological facts owe a great deal to the co-operation of the author's colleagues at Edinburgh University. The original recordings and 400 spectrograms were produced by Miss Shiona Harkness, M.A., now of the College of Arts, Science and Technology, Zaria, Northern Nigeria. Many facts about articulations, almost the whole of the information about air-pressures, the instrumentation for the preliminary work on intensities and much fruitful discussion were generously given by Dr. Peter Ladefoged ; Dr. Frances Ingemann read the manuscript and made several helpful suggestions ; Mr. David Abercrombie, Head of the Phonetics Department, encouraged the work throughout and contributed much information about articulations. Mr. J. Anthony supervised the technical standards. To all these and to many others who have helped as subjects or with illustrations or secretarial assistance, grateful acknowledgement is made.

References

Abercrombie, D. (1956). Direct palatography. *Zeitschr. f. Phonetik,* **10,** 21.

Anthony, J. (1954). New method for investigating tongue positions of consonants. *Science Technologists Association Bulletin,* **4,** 2.

Cooper, F., Delattre, P., Liberman, A., and Gerstman, L. (1952). Some experiments on the perception of synthetic speech sounds. *J. acoust. Soc. Amer.,* **24,** 597.

Draper, M., Ladefoged, P., and Whitteridge, D. (1957). Expiratory muscles involved in speech. *J. Physiol.,* **138,** 17.

Draper, M., Ladefoged, P. and Whitteridge, D. (1959). Respiratory muscles in speech. *J. Speech & Hearing Res.,* **2,** 1.

Fischer-Jørgensen, E. (1954). Acoustic analysis of stop consonants. *Miscellanea Phonetica II,* 42.

Halle, M., Hughes, G. and Radley, J. (1957). Acoustic properties of stop consonants. *J. acoust. Soc. Amer.,* **29,** 107.

Harris, K. (1956). Some acoustic cues for the fricative consonants. (Report from Haskins Laboratories, New York.)

Harris, K. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech,* **1,** 1.

Hughes, G., and Halle, M. (1956). Spectral properties of fricative consonants. *J. acoust. Soc. Amer.,* **28,** 303.

Ladefoged, P. (1957). Use of palatography. *J. Speech & Hearing Disorders,* **22,** 764.

Ladefoged, P., Draper, M. and Whitteridge, D. (1958). Syllables and stress. *Miscellanea Phonetica III,* 1.

Lawrence, W. (1953). The synthesis of signals having a low information rate. *Communication Theory,* ed. W. Jackson (London).

Liberman, A., Delattre, P. and Cooper, F. (1952). The role of selected stimulus variables in the perception of the unvoiced stop consonants. *Amer. J. Psychol.,* **65,** 497.

Liberman, A., Delattre, P., Cooper, F., and Gerstman, L. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychol. Mono.,* no. 379, 1.

Meyer-Eppler, W. (1953). Untersuchungen zur Schallstruktur der stimmhaften und stimmlosen Geräuschlaute. *Zeitschr. f. Phonetik,* **7,** 89.

Schatz, C. (1954). The role of context in the perception of stops. *Language,* **30,** 47.

Strevens, P. (1958a). The performance of "PAT", the six-parameter speech synthesiser designed by W. Lawrence. *Revista do Laboratório de Fonética Experimental, Coimbra,* IV, 5.

Strevens, P. (1958b). Edinburgh's artificial talking machine. *University of Edinburgh Gazette,* **20,** 4.

Strevens, P. and Anthony, J. (1958). The performance of a six-parameter speech synthesizer. *Proc. 8th Intern. Congr. Linguists* (Oslo), 214.

Way, R. (1957). The Articulation of Certain "Alveolar" Plosives. Dissertation for University of Edinburgh Diploma in Phonetics, 1957. (Unpublished).

A THEORY IN DISTRIBUTIONAL SYNTAX: CLASSES AND CONSTANTS

MONICA LASCELLES

University of Tasmania

Developments are made from the definitions of a syntactic class and a syntactic constant given in the article " Fries on Word Classes ", *Language and Speech*, 2 (1959), 86. They are related to work by Zellig S. Harris. The theory is applied to English but it is hoped it will be more generally useful.*

1.0. DIFFERENTIATION BETWEEN SEMANTIC AND SYNTACTIC ELEMENTS

1.01. We begin studies on a corpus where the morphemes only are identified, and have to determine whether all variations in their occurrence are going to be formalized or not. For example, by taking *I'll make the sandwiches with meat* and comparing it with *I'll make the chair with cedar,* it can be found that *chair* will not occur in exactly the same sequence as *sandwich* does. Variations like this one can be dismissed as semantic in favour of other distributional features which we call syntactic for the following reasons.

The syntactic analysis of the occurrence of morphemes begins with relationships between individual morphemes in the sequential order of a single given sentence. Here we hold all morphemes in the sentence constant except one, and test out what other morphemes will fill the position of this one. We may then define a class of morphemes which will fill this particular frame, and say it has the frame as a constant feature. Thus *The boy —s the dog* will characterize a particular list called *x*.

We may then free the morpheme in the position preceding the blank, and see what occurs here in relation to the rest of the original sentence which now includes the morpheme previously omitted. Thus *The — likes the dog.* We may call the resulting class *y*. It is then possible to try to free the position of +*s* and examine all the possibilities of morpheme occurrence here, again in relation to the original sentence. The process is repeated for each morpheme position until the last one is analyzed. In some sentences it may be found that the position cannot be freed for substitution because only the original morpheme will occur there. For example, *to,* in *m v + s to v + o t n + o. (He wants to sweep the room).*

* *I would like to express my appreciation of the benefit I have received from discussion with Mr. F. W. Harwood, University of Tasmania.*

The next step is to re-examine each of the classes now created, and to see what other morphemes will substitute in a given position when any of the individual members of each class are selected and held constant in their positions to make a new sentence with the same sequential order. In this way we may keep enlarging the classes until all possibilities of occurrence in this particular sentence order are exhausted.

Thus if we take a simple arrangement of classes such as the following $/w_{a\ldots d}/x_{a\ldots g}/y_{a\ldots i}/z_{a\ldots e}/$ and select $/w_a/x_f/$ — $/z_i/$as a frame, we may find that $y_{g\ldots n}$ will occur. Then using each of the members of the classes in turn in continual re-examinations of each position, a point may be reached where no more substitutions will either be necessary (since we need only examine commonly found words for some large classes) or will occur. We can now finally number the position of each class in sequence.

1.02. The classes then, are established only for this type of sentence and for no other, and their definitions are sentence long. But in some cases we can later discard the full sentence. For, as knowledge grows, it can be found that a partial sentence is sufficient to determine that one and only one class will occur in a given position.

Examination of other sentences may show similarities of distribution or differences, and compact statements of morpheme occurrence may develop with re-definition, i.e. subdivision or enlargement, of already established classes. For example, an established class may be subdivided because in another sentence it may be found that one or more of its members will occur there exclusively. Syntactic recognition of this must be given at some point. Another class might be arbitrarily enlarged because in some other sentence a different syntactic class might also occur in the same position with it. The two could be grouped together for convenience.

It can be found that certain large classes of morphemes which appear with a small class consisting of just one or two other morphemes in a position of one particular sentence type, will also appear with a class of one or perhaps two other morphemes in another position in another particular sentence type, and so on. Similar results may appear for lists of words. The limited series with which the large lists occur may consist of bound forms of the kind known as inflections. So it can be useful to define the class of the large lists in relation to a selected group of inflections as a whole, and to ignore conditions where only one member of the inflectional range will occur. Such conditions may then become part of subsequently developed formation rules.

However, it is not necessary, though very probably more useful in English, to define a word class by an inflectional range. We could also make certain free forms fixedly diagnostic of a large class consisting of more than one syntactic class, and then put in rules to show where one sub-class occurs. For example we could select *the*, $v + o$ and *to* as free forms which are each capable of defining the word classes which immediately follow them. It may be that we can make an arbitrary choice whether it is more convenient to select bound or free forms to provide such compact statements.

1.03. A class constructed from other classes (or constants) though convenient, does not satisfy the definition of a syntactic class, which was outlined in a previous article (Lascelles, 1959, pp. 89 - 90). Thus a morpheme or group of morphemes is to be considered a syntactic constant or a syntactic class, if the substitution of any other morpheme for the unit in question will break the rules of sentence formation in at least one instance. If it should happen that any member or sub-class of a given class is found to have any unique sentence use, then either or both units are syntactic elements in their own right, and the given class is arbitrarily constructed.

It is obvious that each of the noun, verb and adjective inflections for example, has unique uses. So if an inflection range is accepted as a single class, it is necessary to realize that it is conveniently constructed out of syntactic units. Other classes of a different kind can be established. Again, in word classes we can allow sequences of morphemes of word length to obscure temporarily the syntactic classes of the individual morphemes contained in them. (E.g. word formatives like *-able, -ize* and so on.)

The value of erecting inflection classes (which include $+o$ to denote when no inflection appears) is to have as a record that in all sentences certain lists of word length morphemes will always appear with one member of such a class. Other pieces of syntactic information could also be compactly arranged, and combined with formation rules in order to describe the language. (See Lascelles, 1959, pp. 86 foll. for problems in analyzing a text. This article also supplies a key to the formula symbols used here.)

2.0. Relationships with Harris

2.01. The outlined processes for extracting the syntactic elements from a text would be most easily handled by machines. They offer, of course, a re-statement and development of distributional procedures put forward by Harris (1946, 1951).

There are two notable differences. Harris uses a system of numbering (1951, p. 353 foll.) in only a very limited manner, and its general power is not realized nor discussed in major procedural sections. Again, in the present paper, the possibilities of morpheme substitution are exhausted for each sentence in turn, before proceeding to other sentences, and several sentences are not examined at once as recommended in Harris (1946) and Harris (1951, Appendix to 15.4).

The differences should overcome some objections to his procedures. The Harris idea of approximation techniques leading to general classes causes some oversight about the need to present exact environmental characteristics of some elements. This idea appears to follow from the tendency to approach an ideal (which he elsewhere rejects because it is too impractical, e.g. Harris, 1951, p. 244) to classify morphemes together when they have the same total range of environments (e.g. Harris, 1951, p. 251). It would have been better if he had completely accepted his view that " for each class there are particular sentence positions which can be filled by any member of that class and by those alone " (Harris, 1946, 4.1). The present refusal to use sentences variously for substitution diagnosis should help make the environmental limits more clear.

Harris's problems about whether a class should be defined by long or short environments, can also be resolved by specifying what characteristics are chosen arbitrarily to define it, and what are put into formation rules. (See Harris, 1951, pp. 255-256. It has been found that, in fact, Harris relies very much upon inflections as the defining features.) Some formation rules concerning class use can be more involved than others, for on one occasion a long environment is needed to determine the occurrence of one and only one syntactic class, while on another, a short one is all that is needed.

2.02. An important statement in Harris (1951, p. 7) does not make it particularly clear how to overcome the vast range of semantic combinations of morphemes in order to present syntactic rules. Thus:

> ". . . . x and y [morphemes] are included in the same element A [i.e. class] if the distribution of x relative to the other elements B, C etc., is in some sense the same as the distribution of y. Since this assumes that the other elements B, C etc., are recognized at the time when the definition of A is being determined, this operation can be carried out without some arbitrary point of departure only if it is carried out for all the elements simultaneously."

A clarification is required of what is meant by "in some sense", and by the establishment of all class elements simultaneously. It is hoped that this has been done, and an answer is given to some major points made by McQuown (1952) and Fowler (1952). It appears that a resolution in one case of interdependent levels has been achieved. (Compare Chomsky, 1957, p. 57, footnote 7.)

2.03. Some general objections have been made to Harris's idea that morpheme classes can be established on distributional material alone.[1] But the theory in the present paper should show that although the presentation of syntactic information about classes and constants can vary, the same minimal syntactic units for any two investigators will always result so long as the morphemes of a given text have been identified by both in the same way. There will be no alternative syntax. (These remarks, however, are not to be applied to morpheme sequence analysis, for example possible immediate constituents, where special problems arise which are not discussed here.)

In Harris's system, however, he has himself pointed out that alternative morpheme class analyses are possible (Harris, 1951, p. 2, p 173). For example P and R as defined in Harris (1946, 4.1) are given separate status, though they may both occur where other morphemes occur in important sentences.

Consider the morphemes which substitute for one another in the following when using the type of analysis recommended by Harris (1951, appendix to 15.4).

[1] *Haugen (1951, p. 219, 3.6, foll.); McQuown (1952, pp. 500-1); Fowler (1952, p. 504); Fries (1954, p. 57 foll.); See also (a) Hatcher (1956, p. 236 foll.). Such criticism is here implicit. (b) Firth (1951, p. 82). He suggests that if meaning is excluded, then it follows that the ideas of* concept *and* mind *are also excluded. I have never been able to see the necessity of this and hope this paper suggests otherwise. (c) Hoijer (1954). The discussion "The Cultural Content of Language Material" raises questions which relate to Harris's claims on the syntactic level. (d) Pike (1954, p. 22). He accuses American linguists of "concealing data" of a cultural kind by way of scientific rigidity.*

The boy is	here
” ” ”	near
” ” ”	red
” ” ”	open
” ” ”	fast
” ” ”	Bill

There seems at first sight a traditional or meaning basis for the setting up of P which is outside any of the distributional procedures. Nevertheless, there may be formal reasons for recognizing P which are not pointed out in theory by Harris. (The fact that they do not occur in certain inflectional environments, and also that they have unique distributions in other important sentences may be significant.)

Other sentences where general classes can be set up different to those he accepts are not hard to find. For example, *n*/*a*/*v* overlap (e.g. $v + o$ as in *cut*) before *n*. It is hoped that this paper eliminates these possibilities as statements of the syntax of the language and shows them to be possible constructs upon the syntax.

Because the claim has been made that a non-alternative analysis of a text is possible, we therefore do not agree with those critics who maintain that distributional procedures for discovering classes and constants necessarily create merely an imposition upon the language. (Meaning analysis could be an imposition also. Only a very thorough examination would show whether Bertrand Russell (1940) in his analysis of English structure in “ An Inquiry into Meaning and Truth ” uses syntactic or metalanguage concepts which are fully tenable.)

The theory presented has been subject to some elementary testing but requires machine processing to show whether the claims made for it are correct. If so, then the contention that the distributional syntax of classes and constants relates to the way in which people use the English language, will be proved valid or invalid according to whether the elements taken as the minimal syntactic values or constants, i.e. the morphemes, are units which people do use. Proof would no longer appear to concern distributional method but the nature of the element involved. Thus distribution would certainly appear capable of yielding some results which relate to human response, and to the culture of those using the language.

2.04. The outlined processes seem to be of advantage in clearing up some problems in “ discourse analysis ” where transformation material is neither clearly syntactic nor semantic, and where it is difficult to know just what kind of manipulations are being made upon any text. The writer has not been able to see that the types of analysis put forward by Harris (1946, 3.2 and 1951, appendix to 15.4) cannot allow semantic analyses of morpheme distribution which can compare closely with the results of procedures recommended in “ Discourse Analysis ” (Harris, 1952). If we were to proceed with an analysis of what particular morphemes appear with what other particular morphemes where no morpheme frame is held constant (as in classes 1-6 of the appendix) it appears that a kind of “ discourse analysis ” could result.

Examination of the occurrences of morphemes in varieties of individual sentences

(Harris, 1946, 1951) can allow the features by which the classes are defined to remain unclear. It is true that the frame *They will —* for instance, is picked out in the appendix as one which is common to all the words which fit in a selected variety of other positions (although it will permit another word, *it*, to occur as well). But what the constant features of these other positions are in what types of sentences, remains obscure, and semantic restrictions appear to be included.

In Harris (1946, 3.2 and footnotes) an example of how to begin analysis of morphemes into classes is presented. It is made clear that because *house* and *poem* can both appear in *That's a beautiful —* they are to be put into one class. Even though *poem* will alone appear in *I'm writing a whole poem this time,* the class is to hold generally and not for the particular sentence containing *beautiful,* because *house* alone will appear in a comparable sentence, *I'm wiring a whole house this time.*

> "... morphemes having slightly different distributions are grouped together into one class if the distributional differences between their environments correspond to the distributional differences between the morphemes. That is, if *poem* and *house* differ distributionally only in the fact that *poem* occurs with *write,* and *house* with *wire,* and in comparable differences, we put *poem* in one class with *house,* and simultaneously put *write* in one class with *wire.*" (Harris, 1946, p. 164.)

The essential question is how we know that *write* and *wire* may be called comparable environments for *house* and *poem,* without any further information. If a decision apparently as arbitrary as this is made, then we may also make a great variety of morphemes comparable.

Take: *The boy likes the meal hot*
The boy likes the public garden

Since *meal* and *public* will both occur in *They do not like the —,* it may be said that in the above examples the words provide the same environments for *hot* and *garden.* These last may consequently be classed together (a) because of similarity of environment, and (b) because of their substitution for one another in other environments such as *The — tap is here.*

(Compare the results of analysis in 2.03, where an opposite kind of non-syntactic classification is permitted as the result of not comparing the class established by substitution within one sentence, with that established in another.)

3.0. Some practical advantages

3.01. The semantic - syntactic differentiation has some particular advantages. When we test what fills a position in relation to the class order of a formula, we are testing the syntactic occurrence of morphemes. When we examine a position in relation to any member of one or more classes we are examining the semantic occurrence of morphemes. Thus *t boy — t n + o* enables us to define the distribution of morphemes in relation to the particular value, *boy,* in a particular kind of sentence.

We would also be able to hold that *+s* and *+ed* are in this sentence members of a class. But it is not necessary and it is thought erroneous, to call them semantic members of the class. Both will appear elsewhere as logical constants. For example $+n$ is irreplaceable in $t\,n + s$ *are* $v + ing$ $t\,n + s$ $v + n$ p $t\,n + o$ and in m $v + d$ $t\,n + s$ i *have* $v + n$ *(The children are eating the apples stored in the shed; I paid the men who have worked).*[2] Although an adjective will appear in the same position as the value of $v + n$, it supplies a substitution for two morphemes and not one. The use of $a+o$ to represent the absence of inflection is for purposes of recording convenience, and $+o$ cannot be counted as a morpheme position. This leaves $+n$ as a logical constant.

3.02. Since when a morpheme appears once as a syntactic constant, it is always a syntactic unit, it remains so even when it is replaceable in other sentence forms. Though we have seen that in this last case it can be treated as a value of a variable, it is a value of a special kind. A treatment which accepted it merely as a semantic unit, would overlook the fact that it supplies a meaning value which is indispensable in the over-all working of the language. In other words, it would ignore its particular nature as a grammatical category. For example, *by* and *with* are usually classed as prepositions, but have passive uses which other prepositions do not.

It may be that some positionally dispensable morphemes are so frequent that their semantic value becomes of great importance in the language. But this does not justify their confusion with genuine syntactic terms.

3.03. The definition is of use in those cases when various inflections, or again, different word classes become mutually substitutable in certain frames. For example, let us take *The wounded require treatment* and free $+n$ for substitution. We find then that $+s$ will occur. Then, if we free *wound* in relation to the two member class which results, we can get both nouns and verbs occurring in the same position—*boy, girl, tire, fatigue* and so on.

We do not wish to say that the members of the second class thus set up have the same relation to either $+n$ or $+s$, as values of other classes have to one another when no logical constants are involved. For example take *the tall man, the bright light, the smooth path.* Here we need to show that the differences of occurrence among the values of $a + o$ in relation to those of $n + o$ are of a purely semantic kind.

3.04. The procedure could also deal with distributional features such as the splitting of nouns according to their *he* or *she* substitutions (Harris, 1951, p. 303), and according to their combinations with prepositions (Harris, 1951, p. 312 footnote 14), where it has not always been easy to decide whether the subdivisions are syntactic or semantic.

Finally, the differentiation of syntactic units ought to show whether ambiguous meaning transformations are due to the state of the syntax in the language, or to that of its semantics. Harris's equivalences (1951, p. 272) could be rejected as semantic and outside the scope of a grammatical analysis, while views given by Wells (1947) could be re-presented.

[2] are *and* have, *like* to *in a previous formula, are kept as syntactic constants.*

4.0. Remarks on constants

4.01. Harris gives practically no attention to syntactic constants, although as Fries was well aware (1952), " function words " are of considerable importance. The stress by Fries upon them has been valuable in structural studies.

Fries maintains that the distinctive feature of a function word is that it carries structural meaning within itself, isolated from any formal appendage. But this is not exclusive to function words, for it is true of many words which are members of either the noun or verb class, for example, that they can be listed in isolation as either $n + o$ or $v + o$, and will be recognizable as members of their syntactic classes. For example, *woman, cardigan, spectacle ; think, behave, tolerate.*

Logicians, however, have also tended merely to list the logical signs without being able to solve how they are different from the values of the variables (Quine, 1952, p. xv ; 1941, p. 1 ; 1951, pp. 1, 2 compared with p. 6 ; Reichenbach, 1947, p. 318 foll.; Russell, 1948, pp. 136, 269-70 ; Tarski, 1949, pp. 3-4, 18 ; Carnap, 1942, pp. 56 - 9).

It may be put forward that the essential characteristic of Fries's function words is that they are limited groups of individual free forms for which no other outside words can substitute to create a particular arrangement of syntactic units for a sentence of a given positional length. It may be that in other types of sentences they will not even substitute for one another, but must be listed separately. In this case they would satisfy the definition of a syntactic constant if they were single morphemes.

4.02. A set of function word groups or constants different to the groups Fries establishes is required. Not all the subdivisions of the 154 words he recognizes, for example, in the particular sentence types examined, exclude substitution of their values by words outside of each group. Nevertheless, since it will be found that for various frames various re-groupings will have to be made, or that some words which are at one time special values of a class, will be constants at others, there is at least a good deal of truth in his remark that the function words must be remembered as items. The variables *n, v, a,* etc., on the other hand, allow of a wide variety of substitutions ; that is, they have very large numbers of values which can be selected freely.

Although function words can be put into groups, individual members of a group may occur only with certain lexical items. For example, various prepositions, or the pronouns *who* and *which,* are linked with different meaning ranges of the nouns. The subdivisions of a group which may consequently be created will be called semantic if it is true that when any of the members is used in any given frame, it does not alter the occurrence of the syntactic elements, but relates only to the values of the variables. Thus in

The $n + s$ $v + d$ *the* $n + o$ — $v + s$ $a + ly$

who and *which* will fill the same syntactic frame ; whereas *the* and *some, than* and *as,* must be treated either as four separate constants or values of four separate groups (depending on fuller study) because we find

Some $v + o$ $a + ly$ but not *The* $v + o$ $a + ly$

and *The* $n + o$ *is* $a +$ *er than the* $n + o$

but not *The n + o is a + er as the n + o*

4.03. It is now possible to treat Fries's view that " In the words of our fifteen groups it is usually difficult if not impossible to indicate a lexical meaning apart from the structural meaning which these words signal." (1952, p. 106.)

It may be said that if the words belong to a group then they will at least have the meaning characteristics of the frame in which they are used, and if they are individual constants, then there is no need to try to separate their lexical from their structural meaning. As constants, they may be called autonomous symbols of the metalanguage (Carnap's terminology), whose descriptive or semantic content is automatically carried into the syntax by the necessity of retaining the object language words as syntactic constants (Carnap, 1937). These remarks may be compared with those of Chomsky (1957, pp. 104-5) and should help to clarify the nature of his " so-called ' grammatically functioning ' morphemes " such as *ly, ing,* etc.

5.0. Conclusion

5.01. It appears that there is no way of knowing distributionally that different individual morphemes in a variety of sentences provide the same class frame (i.e. syntax frame) as others, unless we first begin by exhausting their positional relationships to one another in individual sentences. (Compare the great variety of morphemes in a series of six positional sentences.)

For example, it is only when this is done that we can classify morphemes by a feature which is common to all or a great number of sentences. When each kind of sentence is examined, the knowledge can be gained that some morphemes always occur with one of a range of another kind, either always or under conditions which may be stated.

Unless we exhaust the substitution possibilities for each sentence before undertaking their comparison, we may be in danger of setting up semantic restrictions of occurrence which may not be either accurate or useful for a system, perhaps required in a machine, which needs a syntactic structure on the one hand, and freedom for the introduction of special semantic rules on the other. This is one of the main reasons for an insistence upon the distinction between syntax and semantics. Some semantic rules may not be valid in one system (for example one field of knowledge), while they are in another, but the syntax will be valid generally.

5.02. The views expressed here have been influenced by Carnap (1937). A syntactic constant may be compared with a logical constant as defined by him and similarities and differences can be found. It is perhaps of interest that similar methodology in discovering the metalanguage or syntax of artificial and natural languages can be used. Similarly, a syntactic constant has been seen as an autonomous symbol still capable of retaining its descriptive (i.e. semantic) value.

It can be held that in the constants we have a unification of two aspects of a language. In any syntactic system and therefore in an analytic sentence there will be a certain amount of descriptive content which may be studied or ignored according to convenience, and in any semantic system there will be syntactic content which may be treated in the same way.

This last is true because, for example, a descriptive content may be provided for the class variables also, although this does not in any way alter their syntactic nature, nor make the possibility of study of the positional arrangement of the signs of a language, and so of analyticity, impossible to separate from descriptive study.

5.03. While it is agreed with Chomsky (1957, p. 104) that " structural meaning " is often a dubious notion, it is also stressed that there is a necessity to establish more precisely the value and nature of such a notion. The need to know how to proceed in order to extract any instance of it, confuses the answers to both these questions.

When they have been found, the problem of whether we can assign structural meanings to units as explicitly or loosely as semantic meanings are assigned to units, can be resolved. There is clearly more than one meaning to many syntactic units, but the same is true of semantic ones.

It is felt that it is not possible to say yet whether the assignment of structural meanings " is a step of questionable validity " (Chomsky, 1957, p. 104) until we know more about what we mean by " structural meaning ". In fact, Chomsky himself indirectly suggests the need for retaining the idea, when he says that to understand a sentence, we must understand more than one linguistic level (Chomsky, 1957, p. 87).

There is also some disagreement with Chomsky that for " discovery procedures " to be accurate, they most likely must be very complicated, non-rigorous or impractical (Chomsky, 1957, pp. 52-3, 56). It is not accepted that " it is unreasonable to demand of linguistic theory that it provide anything more than a practical evaluation procedure for grammars " (Chomsky, 1957, p. 52).

The theory outlined here suggests a rigorous procedure for discovering minimal syntactic units—that is, constants and variables of single morphemes—which is not too involved for commonly used sentences and for already constructed machines. However, the extent of its value as a series of clear instructions can only be proved by thorough testing.

References

Bloomfield, L. (1933). Language (New York).

Carnap, R. (1937). The Logical Syntax of Language (London).

Carnap, R. (1942). Introduction to Semantics (Cambridge, Mass.).

Chomsky, N. (1957). Syntactic Structures (The Hague).

Firth, J. R. (1951). General linguistics and descriptive grammar. *Trans. Philol. Soc.*, **82**, 82.

Fowler, M. (1952). Review of Z. S. Harris, Methods in Structural Linguistics. *Language*, **28**, 504.

Fries, C. C. (1952). The Structure of English (New York).

Fries, C. C. (1954). Meaning and linguistic analysis. *Language*, **30**, 57.

Harris, Z. S. (1946). From morpheme to utterance. *Language*, **22**, 76.

Harris, Z. S. (1951). Methods in Structural Linguistics (Chicago).

Harris, Z. S. (1952). Discourse analysis. *Language*, **28**, 1.

Harris, Z. S. (1952). Discourse analysis: a sample text. *Language*, **28**, 474.

Harwood, F. W. (1955). Axiomatic syntax. *Language*, **31**, 409.

Hatcher, Anna G. (1956). Syntax and the sentence. *Word*, **12**, 234.

Haugen, E. (1951). Directions in modern linguistics. *Language*, **27**, 211.

Hoijer, H. (1954). Language in Culture (Chicago).

Jespersen, O. (1924). The Philosophy of Grammar (London).

Jespersen, O. (1933). Essentials of English Grammar (London).

Lascelles, Monica (1959). Fries on word classes. *Language and Speech*, **2**, 86.

McQuown, N. A. (1952). Review of Z. S. Harris, Methods in Structural Linguistics. *Language*, **28**, 495.

Pike, K. L. (1954). Language in Relation to a Unified Theory of the Structure of Human Behavior, Part I (Glendale, Calif.).

Quine, W. van O. (1941). Elementary Logic (Boston).

Quine, W. van O. (1951). Mathematical Logic (Cambridge, Mass.).

Quine, W. van O. (1952). Methods of Logic (London).

Reichenbach, H. (1947). Elements of Symbolic Logic (New York).

Russell, Bertrand (1940). An Inquiry into Meaning and Truth (London).

Russell, Bertrand (1948). Human Knowledge. Its Scope and Limits (London).

Tarski, Alfred (1949). Introduction to Logic (New York).

Wells, R. S. (1947). Immediate constituents. *Language*, **23**, 81.

THE CONTROL TOWER LANGUAGE: A CASE STUDY OF A SPECIALIZED LANGUAGE-IN-ACTION*

WILLIAM H. SUMBY
Operational Applications Laboratory, Air Force Cambridge Research Center, Bedford, Massachusetts

A methodology is presented by which the constraint imposed upon a sublanguage by linguistic and non-linguistic factors is estimated. The asymptotic redundancy associated with the control tower language, when the materials analyzed were predicted letter sequences, was estimated to be 75 per cent compared to 55 per cent for newspaper text. The average constraint imposed upon the selection of message units by the physical situation was estimated to be approximately 82 per cent. When the interaction of situational and linguistic constraints is considered, the estimated redundancy for the language-in-action is increased to 95 per cent.

INTRODUCTION

The control tower language is an excellent example of a specialized sub-language. Such language is used in that area where the most critical phases of flight are accomplished, landings and take-offs. Such manœuvres are land-monitored and directed vocally by the control tower system. Because of the importance of control tower operations to the success of the Air Force Mission, the Operational Applications Laboratory was requested to examine the effectiveness and structure of the control tower language. A summary paper has been previously presented (Frick and Sumby, 1952). The present article summarizes the methodology followed in the analysis of this specialized language-in-action.

The principal decisions that had to be made concerned the appropriateness of the levels of analysis to be employed. Primary consideration is given to the method whereby the constraint imposed by the physical situation was estimated.

WORD COUNT

An exploratory and rather easily obtained estimation of the representativeness of a specialized language, relative to its parent language, may be determined by a comparison of word-count samples. A sample of 5,179 words obtained from transcribed control tower messages and a sample of equal size obtained from the editorial pages of newspapers were the materials compared. The word-counts are described in Fig. 1 in terms of the familiar Zipf (1949) analysis : log frequency vs. log rank. The newspaper count, summarized by the thin line, closely parallels the slope of minus one obtained by other investigators (Condon, 1928 ; Zipf, 1935) studying a more

* *This is Technical Note AFCRC TN 59-17 of the Air Force Cambridge Research Center. This research was carried out under Air Research and Development Command Project 7682, Information Processing by the Human Operator.*

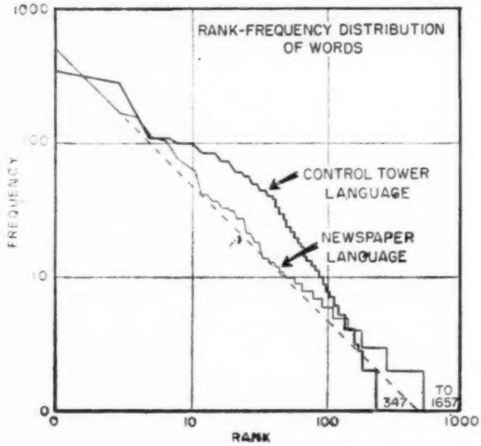


FIG. 1. The frequency of occurrence of the different words in the control tower transmissions and in newspaper text as a function of the rank of the words.

general language. The deviation from this slope by the control tower materials, on the other hand, indicates that a large number of words, relative to the total number of different words, is frequently used, whereas relatively few words are used infrequently.

It is noteworthy that the frequency of numerals constitutes approximately 50 per cent of all words used in the control tower language. Such words were omitted from the word count and are not reflected in the curve. Since numerals here are, for the most part, used as simple identifications and as such are extremely redundant, it seemed more realistic to omit them from the comparison. Their inclusion would have, of course, exaggerated the difference which occurs.

LETTER BY LETTER PREDICTION

The word-counts demonstrated only that the specialized language of Air Force control tower transmissions does not constitute a representative sample of English, using either the Zipf function or the newspaper function as the criterion. It does not afford a quantitative estimate of the information flow. In order to obtain a quantitative estimate of the relative uncertainty of the control tower language, it is necessary to employ a different procedure. The " letter guessing " procedure of Shannon (1949) provides a method for estimating the uncertainty of a set of sequential materials. It is based on the proposition that the predictability of the materials is a reasonable measure of the informational uncertainty of the materials.

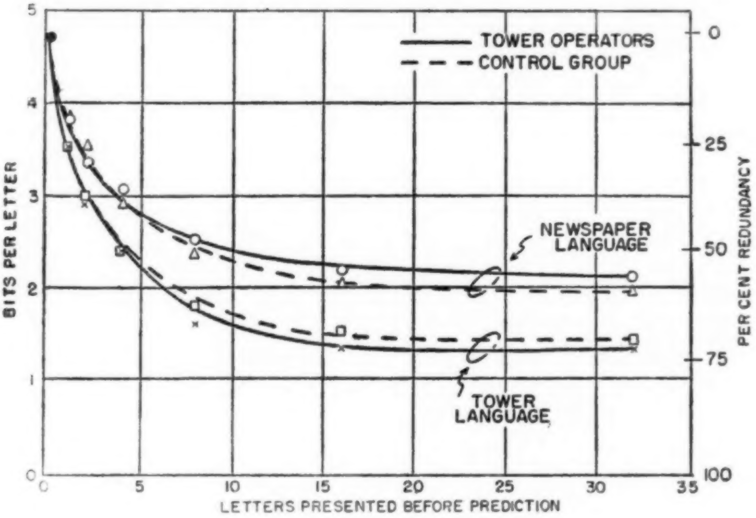


FIG. 2. The uncertainty in bits and the per cent redundancy present in both control tower and newspaper letter sequences plotted as a function of the number of letters contained in the stimulus sequence.

One hundred sequences of newspaper material and of control tower material were randomly selected for each of the following lengths : 1, 2, 3, 5, 9, 17 and 33 letters. The last letter was dropped and the remaining 0, 1, . . . or 32 letters were presented to subjects for prediction of the final omitted letter. Twenty-seven alternative responses (26 letters plus the space) were available for selection. Each subject was tested individually and responses were continued until the correct one was emitted.

Two groups of subjects were used ; a group of six experienced enlisted control tower operators ; and, a corresponding control group of airmen who were unfamiliar with the control tower language.

The results are presented in Fig. 2. The abscissa is the context provided to the subjects, i.e., the number of letters of the materials available before prediction. The left-hand ordinate is the uncertainty of prediction per letter, measured in binary units, or bits. The right-hand ordinate is the redundancy, relative to the maximum possible uncertainty of $\log_2$ 27 or 4.75 bits.

It was noted that the asymptotic redundancy associated with the tower and newspaper materials is almost 75 per cent and 56 per cent respectively. The estimate for the printed newspaper materials is in general agreement with several workers (e.g., Shannon, 1951 ; Black, 1955). Further, the uncertainty changes little after 8 or 16 letters previous context. Finally, only a slight effect of extensive experience with the control tower language is noted ; whereas, the experienced control tower group scored lower than the control group on newspaper text, the experienced control tower group scored higher with the control tower language. The differences between groups, however, are considerably smaller than the differences between languages. Further tests, requiring a short response time, also failed to achieve large differences between groups.

SITUATIONAL CONSTRAINTS

The letter guessing procedure provides a vehicle for examining the detailed letter-by-letter structure of randomly selected segments of languages. Languages, of course, are not employed in this random fashion. Rather, related streams of utterances are more nearly the rule. This stream, too, will be constrained by the stochastic structure present in the language.

In order to examine a more molar aspect of the stream of language sequences it is necessary to employ a more appropriate unit for analysis. The "content element" (Frick and Sumby, 1952, Fritz and Grier, 1954) provided such a unit. The content element is a series of words which refers to a given action or thing. A single content element can be specified in a number of ways, e.g., *gear down and locked, gear in the green, gear checked* all refer to the same action or event. The content elements employed in this study are presented in Table 1. A message would be encoded as follows in terms of Table 1. The message, "*Air Force 1234, Bolling Tower, Clear to enter traffic, runway 10, check base*", would be A,B,C,D,G.

Fifteen control tower situations were described to 110 Air Force pilots. In addition, an appropriate air-to-ground message from the pilot to the control tower operator was included. The subjects were instructed to study each situation and the air-to-ground message, and to predict the control tower operator's ground-to-air response for each. An example of a situation with the appropriate air-to-ground message and tower response follows :

Situation—Visibility and ceiling unlimited. No aircraft aloft in immediate control area.

Message—Bolling Tower, this is Air Force 1234, eight miles south of your station, landing instructions.

A typical response—1234, enter traffic, runway 10, check downwind.

The most striking finding of the tabulation of the content elements is that a population of only 14 different elements was necessary to describe every predicted message for the entire 15 situations. Furthermore, no content element was employed more than once in any particular message. Finally, no message was more than 7 content elements in length. The actual distribution of message lengths, in terms of content elements, is given in Table 2.

TABLE 1
Content Element Code

A—Aircraft Identification
B—Tower Identification
C—Clearance
D—Runway
E—Altimeter
F—Winds
G—Check Points
H—Position
I—Check Equipment
J—Traffic
K—Caution
L—Direction
M—Reception Confirmation
N—Maintain Position

TABLE 1. The coding of the content elements occurring in the control tower messages.

TABLE 2
Content Elements Per Message
(pooled over 15 situations)

Length	Number
1	33
2	297
3	594
4	346
5	198
6	132
7	50
Total	1650

TABLE 2. The number of messages analyzed when arranged according to the length of the message in terms of content units.

First, second and third order informational analyses of the sequential structuring of the content elements were performed. Such analyses supplied estimates of the uncertainty for each position in the message, for each successive pair and each successive triad of positions for each situation.

An example of the content element distribution for a typical situation is presented in Table 3. This situation is hereafter referred to as Situation 1. The frequency of occurrence of the content elements and their distributions according to position in the message are given.

Using the following formula these data were analyzed to determine the average element uncertainty, H(x), associated with each position in the message.

$$H(x) = \log_2 N - \frac{1}{N} \sum_{i=1}^{r} n(x_i)\log_2 n(x_i),$$

TABLE 3

	Position							
Element	1	2	3	4	5	6	7	Totals
A	106	3						109
B	3	59		1				93
C	1	20	37	6	1			65
D		23	52	35				110
E		1	8	28	20	9	2	68
F			4	23	46	12		85
G			9	12	20	37	16	94
H		4						4
Totals	110	110	110	105	87	58	18	598

The frequency of occurrence of the content elements for the described situation arranged according to position in the message.

TABLE 4
Single Unit Analysis

H_{max} = 3.91 Bits

Situation	Unit Position							
	1	2	3	4	5	6	7	Avg.
1	0.25	1.78	1.78	2.34	1.95	1.68	0.73	1.50
2	0.30	1.49	1.69	0.87	0.22	0.00	0.00	0.65
3	0.00	1.76	1.74	1.97	0.73	0.18	0.00	0.91
4	0.23	1.17	1.36	0.83	0.00	0.00	0.00	0.51
5	0.22	1.47	2.29	2.32	1.82	1.84	0.22	1.45
6	0.00	1.73	1.99	2.61	2.24	1.84	0.56	1.57
7	0.98	1.46	1.00	0.00	0.00	0.00	0.00	0.49
8	0.30	1.19	1.60	0.62	0.07	0.00	0.00	0.54
9	0.26	1.49	1.43	1.02	0.07	0.00	0.00	0.61
10	0.21	2.32	2.27	1.68	0.45	0.07	0.00	1.00
11	0.13	1.99	2.24	1.73	1.07	0.51	0.15	1.12
12	0.13	1.88	2.65	2.50	2.39	1.85	0.62	1.72
13	0.34	1.37	1.64	1.15	0.41	0.00	0.00	0.70
14	0.07	2.17	2.81	2.29	1.68	0.69	0.34	1.43
15	0.21	1.19	1.93	1.55	0.67	0.00	0.00	0.79
Avg.	0.24	1.63	1.89	1.57	0.92	0.58	0.17	1.00
	93.8	*58.3*	*51.6*	*59.9*	*76.5*	*85.2*	*95.6*	*74.4*
Pooled avgs.	0.27	2.41	3.66	2.26	1.35	0.75	0.25	1.56
over situations	*93.1*	*38.3*	*6.3*	*42.1*	*65.4*	*80.8*	*93.6*	*59.9*

A first order informational analysis of the occurrence and distribution of content elements. The roman figures represent the average number of bits for each unit position. The italicized figures represent the average relative redundancy for the positions.

TABLE 5
Digram Unit Analysis

$H_{max} = 7.71$ bits

Situation	Unit Position						
	1-2	2-3	3-4	4-5	5-6	6-7	Avg.
1	1.83	2.84	3.06	3.52	2.78	1.92	2.66
2	1.56	2.39	2.19	0.97	0.22	0.00	1.22
3	1.76	2.78	2.91	2.31	0.81	0.18	1.79
4	1.31	1.52	1.54	0.83	0.00	0.00	0.87
5	1.51	2.33	3.36	3.40	2.22	0.51	2.22
6	1.73	2.62	3.42	3.92	3.37	2.05	2.85
7	1.52	1.55	1.00	0.00	0.00	0.00	0.68
8	1.24	2.02	1.95	0.66	0.00	0.00	0.98
9	1.53	2.16	1.76	1.07	0.07	0.00	1.10
10	2.37	3.75	3.29	1.84	0.45	0.07	1.96
11	1.99	3.38	3.10	2.15	1.17	0.53	2.05
12	1.88	3.56	4.22	3.80	3.36	2.15	3.16
13	1.44	2.17	2.21	1.31	0.41	0.00	1.26
14	2.17	3.97	4.03	3.06	1.94	0.72	2.65
15	1.22	2.37	2.77	1.82	0.67	0.00	1.48
Avg.	1.67	2.63	2.72	2.04	1.16	0.54	1.80
	78.4	*65.9*	*64.7*	*73.6*	*85.0*	*93.0*	*76.7*
Pooled avgs.	2.48	4.64	4.46	2.97	1.69	0.85	2.85
over situations	*67.9*	*39.9*	*42.2*	*61.5*	*78.1*	*89.0*	*63.1*

A second order analysis.

where N is the total number of cases, x_i the particular values of x, and $n(x_i)$ is the number of occurrences of x_i in a sample of N events. For purposes of the analysis it was assumed that each message was 7 content elements in length. When the message was less than 7 elements in length, a space element was included in each vacant position. This procedure tends to overestimate the uncertainty somewhat, although the difference is extremely slight for the first five positions for most situations. The results of the first order, or single unit, analysis are given in Table 4. The data presented in Table 3 were analyzed to obtain the results in Table 4, Situation 1. The uncertainty associated with each position for each situation is presented.

The summarizing italicized figures are the relative redundancy measures, relative to H_{max} or 3.91 bits, determined in the following manner :

$$R = \frac{H_{max} - H(x)}{H_{max}}$$

These figures were simply totalled and divided by the appropriate N to determine the row and column marked " Avg. " The row " pooled averages over situations "

TABLE 6
Trigram Unit Analysis

H_{max} = 11.41 Bits

Situation	Unit Position					
	1-2-3	2-3-4	3-4-5	4-5-6	5-6-7	Avg.
1	2.88	3.79	4.00	3.66	2.17	3.30
2	2.46	2.40	1.01	0.22	0.00	1.22
3	2.78	3.50	2.59	0.87	0.00	1.95
4	1.57	1.91	0.83	0.00	0.00	0.86
5	2.33	3.58	3.59	2.31	0.93	2.55
6	2.62	3.89	4.13	3.83	2.19	3.33
7	1.59	1.10	0.00	0.00	0.00	0.54
8	2.05	2.05	0.67	0.07	0.00	0.97
9	2.16	2.22	1.13	0.07	0.00	1.12
10	3.79	4.09	2.09	0.45	0.07	2.10
11	3.38	3.76	2.34	1.26	0.61	2.27
12	3.56	4.70	4.64	3.72	2.68	3.86
13	2.17	2.38	1.36	0.41	0.00	1.26
14	3.97	4.55	3.47	2.28	0.84	3.02
15	2.37	3.05	2.09	0.70	0.00	1.64
Avg.	2.65	3.13	2.26	1.32	0.63	2.00
	76.8	*72.6*	*80.2*	*88.4*	*94.5*	*82.5*
Pooled avgs.	4.69	5.32	3.54	1.98	0.97	3.30
over situations	*58.9*	*53.4*	*69.0*	*82.7*	*91.5*	*71.1*

A third order analysis.

shows the uncertainty associated with the element position regardless of the situation.

The same procedure was carried out with successive pairs and triads of content elements. The results of these analyses are presented in Tables 5 and 6.

The H_{max} reported on each of the tables is equal to :

$$\log_2 \frac{N\,!}{(N - n)\,!}$$

It is shown in Tables 4, 5 and 6 that the average relative redundancy associated with single element, diagram and trigram position is 74.4, 76.7 and 82.5 per cent respectively. In addition to the constraint imposed upon the selection of content elements by the situation at hand, we find that a probabilistic structure of such elements is evidenced. That is, we find additional constraint imposed by the use of a content element upon the selection of the succeeding elements. However, there is an increase in the relative constraint of only 8 per cent when the item of analysis is the trigram unit and compared with the single unit. At the third order of analysis, then, the average constraint imposed by the situation on the selection of content elements is approximately 82 per cent.

It is interesting to note that when all of the content element responses were pooled over situations, the average constraint imposed upon the selection of elements for each position is approximately 60 per cent for single elements and 70 per cent for triads of elements. This indicates that when a particular content unit is used in a message, the probability of other specific units following it is high, regardless of the situation.

Discussion

Control tower language is almost completely unintelligible to the neophyte, a condition apparently attributable, at first glance, to the wide-band masking noise present. However, after extensive experience with the *overall* system the elements of a message can be easily specified. It appears, therefore, that knowledge of the typical situational procedures is the most constraining factor of this sublanguage.

The pilot's knowledge of the situation and the procedures followed in the situation reduce the gross uncertainty of the message approximately 82 per cent. Since the messages are also subject to the linguistic constraints, previously estimated, the residual uncertainty of 18 per cent, in terms of letter sequences, is further reduced by 72 per cent, or 13 per cent of the total uncertainty. The estimated average relative redundancy is 95 per cent with respect to what could have been transmitted. This procedure, of course, assumes the independence of the linguistic and situational constraints. If such an assumption is questionable, a more conservative estimate may be obtained by assuming the newspaper approximation of redundancy to be a more realistic measure of actual linguistic constraint. In this case the residual uncertainty is reduced by 56 per cent or 10 per cent of the total uncertainty, yielding an approximation of 92 per cent relative redundancy.

The primary objective of this paper has been to demonstrate a methodology whereby the situational constraints imposed upon a language might be quantifiably estimated. The series of analyses has revealed that the control tower sublanguage is not a representative sample of the parent language, and also has quantified two sources of constraint operating upon the language at a particular time ; linguistic and situational. It is somewhat startling to discover that so little information is transmitted by the verbal message *per se*.

The high degree of linguistic constraint is evidence that the same messages frequently recur, which in turn is evidence of a highly routine system. Because of such evidence and because the content elements were found to be highly predictable, it appears that the situation serves to delimit severely the elements that can occur. The limitations are restricting to the extent that the receiver expectancy of message occurrence is almost certainty. The verbal message, then, serves, for the most part, only to confirm or refute the expectancy of the receiver.

Obviously the constraint imposed by the situation on this sublanguage is much greater than one would expect to find in the parent language or even other sublanguages. The results do suggest the possibility, however, that the estimates of redundancy reported by other investigators for the general language would have been higher if this source had been considered. Unfortunately, the more general language might not be readily amenable to the specific situational analysis outlined here in that in most cases content elements cannot be specified easily. However, it would seem that the predicted responses to a given verbal sequence in a particular situation would be, at least, correlated and probably lend themselves to element categorization.

References

Black, J. W. (1955). The prediction of the words of varied materials. U.S. School of Avia. Med., Pensacola, Fla., Joint Pro. Rpt. 57.

Condon, E. U. (1928). Statistics of vocabulary. *Science*, 67, 300.

Frick, F. C. and Sumby, W. H. (1952). Control tower language, *J. acoust. Soc. Amer.*, 24, 595.

Fritz, E. L. and Grier, G. W. Jr. (1954). Empirical entropy: a study of information flow in air traffic control. Control Sys. Lab., University of Illinois, Rpt. R-54.

Shannon, C. E. (1949). The Mathematical Theory of Communication (Urbana, Ill.).

Zipf, G. K. (1949). Human Behavior and the Principle of Least Effort (Cambridge, Mass.).

THE PERCEPTION OF ENGLISH STOPS BY SPEAKERS OF ENGLISH, SPANISH. HUNGARIAN, AND THAI: A TAPE-CUTTING EXPERIMENT*

JOHN LOTZ,** ARTHUR S. ABRAMSON,** LOUIS J. GERSTMAN,*** FRANCES INGEMANN,**** AND WILLIAM J. NEMSER**
Haskins Laboratories, New York

American English stops, including *residual* stops, *i.e.*, stops in /s/-clusters after the removal of the /s/, were presented in front of stressed vowels for identification on the one hand to native speakers of American English, on the other, to native speakers of Puerto Rican Spanish, Hungarian, and Thai, languages with differences in the phonetic composition of their stop phonemes. Speakers of American English identified the residual stops with the voiced (lenis) stop ; the others, with the voiceless stop. The results suggest that there is a hierarchic organization among the features of these stops : the lack of aspiration tends to force the evaluation of stops in the direction of /b,d,g/ in American English, whereas in the languages where other distinctions exist, the evaluation is different.

This paper presents some data on the interpretation of speech sounds by speakers of different languages, and the evaluation of the relative importance of cues present in the acoustical stimulus within the framework of the phonemic system of each language. Specifically, the investigation deals with reactions to a set of stop consonants on the one hand by native speakers of American English, on the other hand by native speakers of Puerto Rican Spanish, Hungarian, and Thai, languages with differences in the distinctions among their stop phonemes. In particular, the evaluation of *residual* stops, *i.e.*, American English stops preceding a stressed vowel in /s/-clusters after the removal of the fricative, was studied.

RESPONSES OF SPEAKERS OF AMERICAN ENGLISH

The project originated with a problem in English phonemics. (The linguistic evaluation of the results will not be presented here.) As is well known, before an initial stressed vowel, there occur a voiceless aspirated *fortis* stop and a voiced unaspirated *lenis* stop ; after /s/ only one kind of stop occurs in initial clusters, an unaspirated voiceless stop, as in *spill*. Now, if the /s/ were not there, would native

* *A contributed paper read before the Fifty-Eighth Meeting of the Acoustical Society of America, Cleveland, Ohio, October 22-24, 1959. The research was supported by the U.S. Department of Defense under Contract DA18-119-sc-260.*

** *Also at Columbia University.*

*** *Now at Bell Telephone Laboratories.*

**** *Also at the University of Kansas.*

speakers of American English associate and identify this stop with the initial aspirated stop or with the voiced stop in a forced-choice situation ? The stop after /s/ differs from both, from the /p/ in the lack of aspiration and from the /b/ in the lack of voicing, whatever other cues there are.

The following experiment was devised to investigate this problem. A set of 18 monosyllabic words was prepared, chosen so that all single stops and stops in /s/-clusters occurred, and these stops and clusters occurred before one front and one back vowel. The following matrix sums up the stimuli:

pill	*till*	*kill*
bill	*dill*	*gill*
spill	*still*	*skill*
pore	*tore*	*core*
bore	*door*	*gore*
spore	*store*	*score*

The words were recorded by three native speakers of American English. Then the initial friction in the words beginning with /s/ was removed, and a randomized tape was made containing the remaining portions and the other words. The stops occurring in the original /s/-clusters, thus mutilated, will be called *residual stops,* symbolized in the figures by (s). The stops were tested in the context of words rather than in isolation, so that they could be identified as parts of meaningful utterances instead of isolated segments. The tape was offered to 35 native speakers of American English, who were asked to identify each stimulus with one of the 12 words having initial stops.

Fig. 1 shows the format in which we have tabulated the responses. The labels at the bottom indicate our interpretation of the results. A check of the tape after the test revealed that one *score* and one *spill* were defectively recorded.

The results, which had been anticipated, can be interpreted in the following way. Initial aspiration is a stronger cue for fortis stops in English than lack of voicing, hence the residual stops—lacking both aspiration and voicing—were identified with voiced lenis stops, rather than with the aspirated fortis ones.

One aspect of the mutilation of the /s/-clusters, however, must be justified. Removal of the initial friction also obliterates the bounded silence between the end of the friction and the release of the stop, and we know from other experiments that a silent stretch between speech sounds can affect the judgment in the opposition among stops. That this is not the case in our experiment can be inferred from the following considerations. (Because the residual stops were identified with voiced stops, we have to show only that the distortion did not alter them in the direction of voiced rather than aspirated stops.):

1. There is no evidence that silence produces the effect of aspiration. Experiments have shown that the duration of a gap has a significance for the *fortis/lenis* (or *unvoiced/voiced*) opposition of stops in intervocalic position, but not for *voice* versus *aspiration.*

2. The aspiration is clearly sequential to the closure in time ; in reversed speech the aspiration is heard as preceding the stop closure. It is improbable, therefore, that

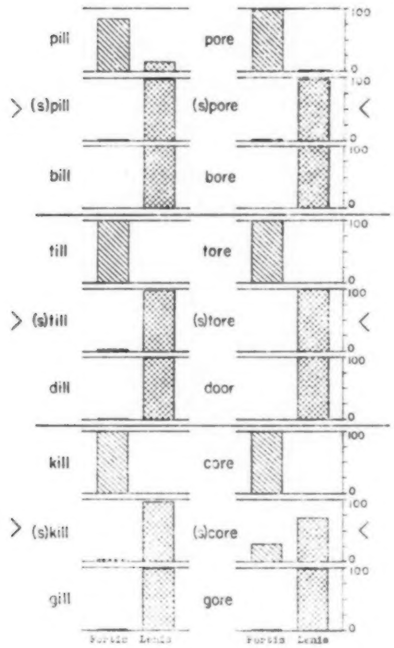


Fig. 1. Per cent responses of American subjects.

a sound feature that precedes the stop release in normal speech would produce the effect of aspiration.

3. Moving the frictional part, corresponding to /s/, back and forth along the time axis with reference to the explosion does not produce any effect of aspiration. This is the most decisive argument. Increasing the gap by a factor of four dissociates the friction from the rest, which is then heard with a voiced initial stop. When the distance is decreased, the stop effect either remains unaltered or it disappears.

Responses of speakers of other languages

The second part of the experiment deals with the reactions of native speakers of Puerto Rican Spanish, Hungarian, and Thai to these stimuli. The two poorly recorded versions of *spill* and *score* were re-recorded and inserted in the tape ; otherwise, the tape submitted for judgment was identical with the one used for the American listeners. Here the experimental situation is different from the first one. The subjects were asked to identify the stimuli with stops in their native phonemic systems, and of

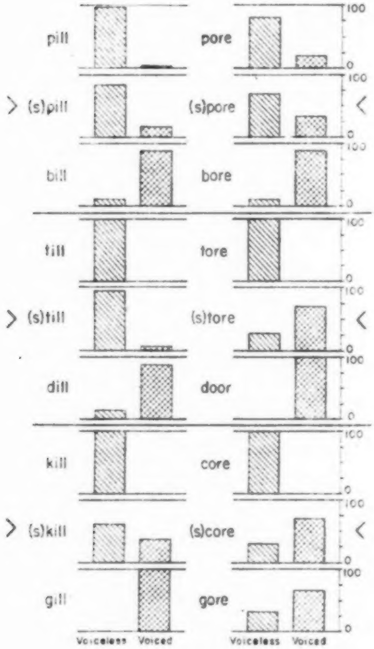


Fig. 2. Per cent responses of Puerto Rican subjects.

course the words were not necessarily meaningful for them. This set of experiments also seeks an answer to the question whether voicing is present in the residual stops, which might have led the American listeners to identify these stops with /b, d, g/, since in these three languages the phonetic feature of the distinction is clearly *voicing* versus *lack of voicing*.

1. *Puerto Rican Spanish*

In Puerto Rican Spanish, stops are distinguished on the basis of voicing with no aspiration present. The responses of the 12 Puerto Ricans are shown in Fig. 2. As can be seen, the trend here is to identify the English aspirated stops as voiceless and the English voiced stops as voiced. The residual stops were heard as voiceless in four out of six cases. Before the back vowels, however, the stops produced with the tongue (apical and dorsal) were predominantly judged as voiced, the labial stop as voiceless. This seems to indicate that the two extremes, aspirated voicelessness and voice, are clearly judged in terms of the speakers' own phonemic opposition, while the residual stops were judged to be, on the whole, voiceless.

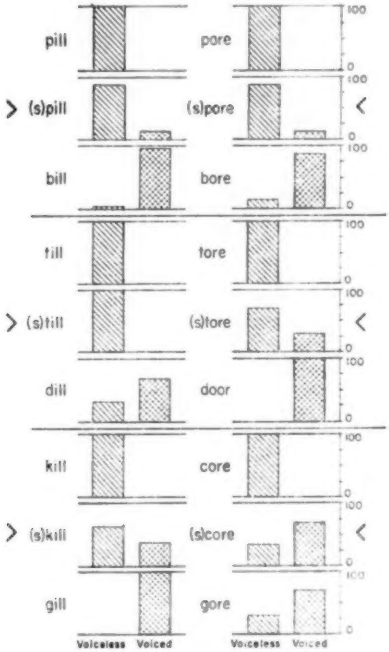


Fig. 3. Per cent responses of Hungarian subjects.

2. *Hungarian*

Hungarian, like Spanish, has an opposition between unvoiced and voiced stops, with no strong aspiration present. Five subjects took the test, one four times and one twice, so the total number of tests was nine. (The repeaters' internal consistency was high.) Fig. 3 shows their responses. Here the results for the residual stops are more clear-cut. They were by and large called voiceless. Only the residual velar stop before the back vowel was largely evaluated as voiced. There is also a definite trend to identify the aspiration as voiceless and the voice as voiced.

3. *Thai*

In the Thai phonemic system there is a three-way opposition among voiceless aspirated, voiceless unaspirated and voiced stops and, in addition, a distinction among three places of articulation: labial, apical, and dorsal. The voiced dorsal is missing. Thus, the phonemic system of stops is:

	LABIAL	APICAL	DORSAL
Aspirated-voiceless	/ph/	/th/	/kh/
Unaspirated-voiceless	/p/	/t/	/k/
Unaspirated-voiced	/b/	/d/	

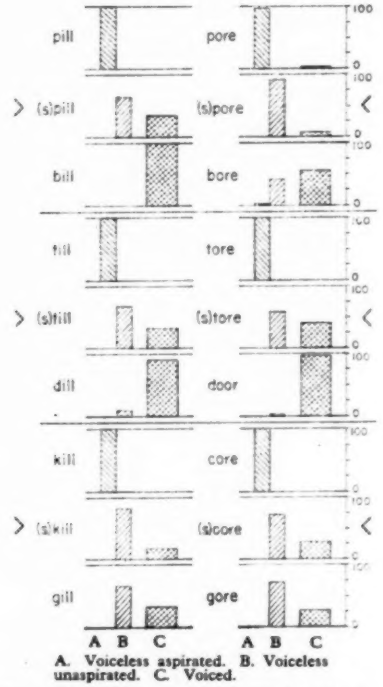


Fig. 4. Per cent responses of Thai subjects.

Twelve subjects took the test, three of them twice, giving 15 scores. The subjects were instructed that they could use the letter *g* for the /ɡ/ sound if they wished. The results are shown in Fig. 4. Aspirated voiceless stops were identified as aspirated voiceless stops in the Thai system, the residual stops for the most part as voiceless unaspirated stops, and the voiced stops as voiced stops. There was considerable confusion about the /ɡ/, which does not occur in the lattice of the Thai phonological system of stops.

Conclusions

The results are less clear-cut for the foreign than for the American listeners. This is attributable to differences in phonetic make-up between the stimuli and the related sounds in the various languages, differences for example in the onset and degree of voicing, the release for the aspiration, and the force of the explosion. However, aspiration for listeners whose languages lack this feature, contributes to the evaluation of these sounds as voiceless. Vowel quality also has some effect on the judgement; back vowels tend in some cases to make the consonants sound voiced. In order to get more complete data, one would have to record the stops of these languages and investigate their evaluation by other speakers, including those of American English.[1]

The results indicate that there is a hierarchy among the cues in the acoustic stimulus for the perception of these sounds in various languages. For American English, the lack of aspiration is a dominant cue for forcing the evaluation of the stops in the direction of /b,d,g/, whereas in the languages where other distinctions exist, the evaluation is different. Our modest data exemplify the reflection of the linguistic categories of the listener's native language in his interpretation of speech sounds.

Our hope is that this kind of study will contribute to the understanding of the relative importance for the perception of speech of various cues present in the acoustic stimulus as well as an understanding of the role of the linguistic system in the evaluation of stimuli. Such studies might also contribute information that will be useful in the teaching of languages.

[1] *A remark about the identification of the place of articulation in the above four sets of experiments is in order. There was practically no confusion in this respect, except by some Hungarian speakers, who identified the labial stops of* pill *and* spill *with their /t/.*

LINGUISTIC PHILOSOPHY IN THE LIGHT OF MODERN LINGUISTICS

GUSTAV HERDAN
University of Bristol

Both linguistic philosophers and their opponents are apparently of the opinion that linguistic structure can be recognised intuitively and that the matter is still virgin ground. However, as this paper shows, much of the confusion and acrimony of the recent dispute on linguistic philosophy is due to the neglect of scientific linguistics as we know it to-day, and to the difference between what Wittgenstein called " linguistics " and what is now understood by that term. In particular, it is shown that two of the so-called " pillars " of linguistic philosophy, viz., the argument from paradigma cases and the argument from the contrast theory of meaning, not only receive no support from, but are invalidated by, scientific linguistics.

In the recent controversy on linguistic philosophy, one participant wondered whether the violence (and, no doubt also the muddle) of the dispute was due to *too much* linguistic analysis (Howes, 1960). Taking this to be a genuine, not only a rhetorical question, the answer is: " By no means. Both the muddle and the acrimony of the dispute are due to *too little* linguistics, i.e., scientific linguistics as we know it to-day."

Neither of the parties to the controversy seems to be aware of the change which the meaning of the term ' linguistics ' has undergone since the origin of linguistic philosophy some decades ago, and in particular since Wittgenstein. This must be regarded as a serious defect of all such discussions because linguistics is to-day something very different from what the parties to the dispute apparently take it to be. What they call ' linguistics ' is simply an anachronism. In modern structural linguistics, the parties to the dispute would not only find the scientific methods required for dealing with the problems they are worrying about, but especially that certain fundamental questions which they still regard as problematic have already been satisfactorily answered.

The consequence of the neglect of the development of linguistic science on the part of linguistic philosophy is that insofar as the latter claims to give information about language structure, it is little more than an amateurish attempt at scientific linguistics, trying to obtain by mere intuition what can only be achieved by the legitimate methods of modern science. Nowhere is this more apparent than in what linguistic philosophy calls ' Language Games ', a concept which is quite fundamental for linguistic philosophy. ' Language Game ' is the term used by linguistic philosophy for what we would describe today as language as a coding system, and therefore as a branch of semiology. In linguistics this conception of language can be traced back to F. de Saussure, a Swiss linguist of the 19th century; it was extended to the smallest linguistic units with distinctive function, the phonemes, by the Russian linguist N. S. Trubetzkoy whose teaching is known as ' phonology ', and to grammatical substitution by the American linguist, Z. Harris, as the Theory of Distribution. The quantification of de Saussure's idea of language as a system of coding is known under the name of Information Theory (C. E. Shannon, 1947) which again presents a great step forward in our understanding of language structure.

All this development is neglected by the linguistic philosophers who depend upon nothing but the intuition about how people use words. How utterly inadequate this is will be shown below.

We shall consider two of what Professor Gellner (1960) has called the four pillars of linguistic philosophy: the argument from paradigma cases (APC) and the contrast theory of meaning.

1. The argument from paradigma cases

It maintains that words mean what they say—though I should not like to say what this means—and that they get their meaning from how they are used by the members of the speech community. For instance, it is maintained by a linguistic philosopher that the freedom of the will can be inferred from the fact that the bridegroom marrying the girl of his choice will smile. Thus, the only-too-willing bridegroom is the paradigma case for the freedom of the will (Flew, 1956, quoted by Gellner, 1960).

It is admitted that the conclusion that something must be true because it is generally said by people was criticised by linguistic philosophers themselves (e.g., Urmson, 1960)—though, as Gellner points out, few things illustrate the silliness of linguistic philosophy better than the fact that there was even the need for Urmson's article to be written—and it is not to this aspect of APC that structural linguistics has anything to say. To dispose of this aspect may safely be left to commonsense only. But structural linguistics has something important to say about the implicitly assumed fact of APC, namely, that there is such a thing as a general use of particular combinations of specified words among the members of a speech community.

Stability in the use of linguistic forms, vocabulary, and grammar, is an important subject of statistical linguistics. As pertinent results may be mentioned that the relative frequency of certain grammar words is, by and large, stable in large samples of the spoken or written language ; that within a given text there is a high degree of stability of vocabulary according to frequency of occurrence of the particular vocabulary items ; that the distribution of word length in terms of phoneme-, letter-, syllable-number is very stable, and so forth (Herdan, 1956).

But nobody, to the writer's knowledge, has so far maintained on the basis of experimental findings by a word count, that the use of certain words in a particular combination is a stable or general phenomenon among the members of the speech community. And this is precisely what is implied in APC, e.g., in our illustration for the three words ' will ', ' is ', ' free '. On the contrary, one of the most surprising results of statistical linguistics is the constancy of certain statistical parameters for different parts of a text, or for different texts belonging to one universe of discourse, *in spite* not only of a lack of identity or even similarity of word combinations, but even of the particular vocabulary items which are used in these combinations ; such parameters are, for example, Yule's Characteristic, Shannon's Entropy (Herdan, 1956).

If, therefore, the linguistic philosopher still wants to uphold the APC pillar, he must adopt the modern methods of statistical linguistics for ascertaining those facts which he has so far silently assumed, viz., a general use of word combinations among the members of a speech community. Most likely he will not be successful in his efforts. Returning to our illustration about the smiling bridegroom, he may perhaps find the number of sentences in which the bridegroom is described as not being free in his decision, to be equal to the number of those in which he is described as having acted entirely by free choice. This is such a commonplace that it seems difficult to believe it should ever become a matter of philosophical discussion. But even if to-day all bridegrooms were smiling, what about the brides of the Victorian era, to judge by the novels we have of those days ? Would the inference be that the universal law of the free will was not functioning under Queen Victoria ? This is strongly reminiscent of Samuel Butler's remark about " God's universal law that man should keep his wife in awe " as not functioning in the case of Mr. and Mrs. So-and-so.

2. The argument from the contrast theory of meaning

A term and its denial between them exhaust the universe, or at least a universe, of discourse. The demarcation line between a concept and its denial may shift as we change the definition of the concept, which often happens. But there is one kind of shift of definition or meaning which according to the linguistic philosopher is both disastrous and characteristically philosophical, and that is to make the criteria for what falls under a concept either so severe, or so loose, that either nothing at all can, or everything must, fall under it.

The concept, if extended so as to comprise everything in the universe, or nothing, is then used without ' antithesis '. People who commit the fallacy of using a term in this way do it, it appears, from the philosophical desire to say something wholly all-embracing, but fail to realise that this ambition is incompatible with saying anything at all.

This argument either intentionally identifies or unintentionally confuses, logic and words. That a concept and its denial exhaust the universe is a law of logic, viz., the principle of contradiction which, as Boole has shown, becomes in the algebra of symbolic logic the ***law of duality*** (Boole, 1854):

$$x = x^2$$

from which we have immediately

$$x(1 - x) = 0$$

which is the algebraic expression of the logical principle of contradiction. For let us give the symbol x the particular interpretation ' men ', and let 0 and 1 stand for ' nothing ' and ' universe ' respectively, then 1 – x will represent " everything in the universe which is 'not men' or the class of 'not men'." Now the product of the expression of two classes represents that class of individuals which is common to both. Hence x(1 – x) will represent the class whose members are at once ' men ' and ' not men ', and the second equation thus expresses the principle that a class whose members are at once ' men ' and ' not men ' does not exist. In other words, it is impossible for the same individual to be at the same time a man and not a man. Now let the meaning of the symbol x be extended from representing 'men' to that of any class of beings characterised by the possession of any quality whatever; the second equation will then express that it is impossible for a being to possess any quality and not to possess it at the same time.

But this is identically what philosophers have called the *Principle of Contradiction*, which affirms that it is impossible for any being to possess a quality and at the same time not to possess it.

However, the linguistic philosopher may rest assured, the duality between a concept and its denial can never disappear. Since it is a fundamental law of thought, thinking is not possible without it, and no old-fashioned philosopher, whatever the faults of his system, could have ever done without it. Else his system would just be nonsense, and even the linguistic philosophers do not accuse the philosophies of Plato, Spinoza, Descartes, Kant and Schopenhauer of this.

Thus, it cannot be the principle of duality for *concepts* used in logic which the linguistic philosopher accuses the classical philosopher of having acted against, because it is simply inescapable, but it must be the extension of a *linguistic* term to the exclusion of the opposite that he has in mind. Now, in language we do not really work with contradictory oppositions. In philosophical systems we do not find that the philosopher speaks of Matter—Not Matter, Spirit—Not Spirit, Good—Not Good, Will—Not Will, but of Matter and Spirit, or Mind and Matter, Good and Evil, God and the Devil, Will and Idea. In general he works not with the *contradictory* opposition of logic, but with the *contrary* opposition, that is, that between the extremes of a series of terms.

It shall not be denied that philosophers are in the habit of extending the definition of a concept beyond its usual or conventional limits—which, *pace* linguistic philosophy, is of course the virtue of all science and philosophy—but, and this is a most important point which the linguistic philosopher has overlooked, the matter does not rest there *because language, like logic, has its limiting or controlling device in the shape of a law of duality.* This is what I have called the principle of linguistic duality to distinguish it from logical duality, which is of the nature of geometrical duality (Herdan, 1960).

One of the general principles which runs through higher mathematics is the principle of duality. It is usually first encountered in geometry, and can there be stated as follows: all the propositions of plane projective geometry occur in dual pairs which are such that from either proposition of a particular pair another can be immediately inferred by interchanging the parts played by the words *point* and *line*. As a specimen of dual propositions we give the following pair :

1. Two distinct points are on one, and only one, line.
2. Two distinct lines are on one, and only one, point.

Thus, if we have any plane figure consisting of points and lines, there exists also a figure consisting of lines and points such that to any point of the first figure lying on a line of that figure there corresponds a line of the second figure passing through a point of this figure which corresponds to the line of the first figure, and so on.

The physical means of sound and symbol are highly incommensurate with the manifold of objects which is to be expressed by them. Only in the primitive stage of linguistic development when the shortcoming of language symbols in this respect is not recognised, or has not yet made itself felt, are languages made so as to duplicate the manifold of objects. Hence the onomatopoetic words and gesture imitations by the organs of speech. Such imitating methods soon prove inadequate and, as the language develops, the sounds and signs acquire greater and greater independence from the denoted object, until the state is reached when sounds and symbols have distinctive function, that is they are means for making distinctions between words, but are no longer linked with meaning.

Since the expression of the content of our consciousness can only be in terms of code symbols, whose relations are subject to laws of language which are not identical with the laws of nature, a certain amount of discrepancy between content and expression is unavoidable. For the sustained description of a situation or a system of relations, we must choose a viewpoint. This introduces a certain arbitrariness in the description which, if carried to the bitter end, cannot but give a distorted picture of reality. The remedy is to change the viewpoint in the manner indicated by geometrical duality. In this way geometrical duality provides the limiting factor of the error element, and with it successive approximations to truth.

The similarity between the law of duality in logic and linguistic duality as used in philosophical systems was stressed by Boole. He says :

" In . . . intimate alliance with that law of thought which is expressed by an equation of the second degree, and which has been termed in this treatise the law of duality, stands the tendency of ancient thought to those forms of philosophical speculation which are known under the name of dualism. . . . The logical elements which underlie all these speculations, and from which they appear to borrow at least their form, it would be easy to trace in the outlines of more modern systems, more especially in that association of the doctrine of absolute unity with the distinction of the ego and non-ego as the type of Nature, which forms the basis of the philosophy of Hegel . . . The wide prevalence of the particular theories which we have considered, together with their manifest analogy with the expressed laws of thought, may justly be conceived to indicate a connexion between the two systems . . . The operation of that law of thought termed in this work the law of duality may have its own peculiar tendency to error, exalting mere want of agreement into contrariety, and thus form a world which we necessarily view as formed of parts supplemental to each other, framing the conception of a world fundamentally divided by opposing powers."

There is, thus, no ' loss of antithesis ' as anticipated by Wittgenstein, because the principle of duality counteracts it. Nowhere can this be better seen than in the apparent difficulties which modern theoretical physics is having with its terminology (Wilkinson, 1959). As the concepts whose meaning is so widely extended as to comprise everything in the universe, approach more and more the nature of idealised geometrical entities, such as points, lines, waves, the description of the universe in terms of language becomes avowedly one of dualities: as atomic physics vs. wave theory, as deterministic vs. statistical physics, as empirical vs. aprioristic physics. As regards the notorious particle-wave duality, the physicist rightly objects to a question such as " But *is* it a wave or *is* it a particle ? " We can only say that from some points of view it behaves like a particle, and from others it behaves like a wave. To regard this as a worry misses the point. We must frankly accept the duality of description. But more: it is only through considering both aspects that our description of ultimate events in the physical universe becomes complete.

To summarize : Linguistic philosophy as such can be justified as the culmination of a long philosophical tradition, namely of critical philosophy as initiated by Kant, whose principal idea is that we must understand our tools of thought—not *before* we use them, as Gellner assumes, but in order to know what to expect in the way of philosophical results. The novelty in linguistic philosophy in this respect is that whereas in Kant's philosophy the tool under examination was the mind, it is now language.

However, in order to do this, linguistic philosophers ought not to assume, as they have done so far, that language structure can be recognised intuitively, but should acquaint themselves with both the methods and the results of statistical linguistics and of statistical duality, to name only these two branches of mathematical linguistics, or structural linguistics as a modern science in the strict sense of the term.

As this paper shows, this necessarily leads to abandoning the unscientific assumptions implied in APC, and to the realisation that the philosophers' use of terms with widely extended meaning is subject to the control principle of linguistic duality, completely comparable to Boole's law of duality, from which it follows that Wittgenstein's fear of the ' loss of antithesis ' in the case of using widely extended terms was unfounded. Not only is the antithesis of logical duality not lost—because it is absolutely indispensible to every kind of thought—but language has created its own ' brand ' of duality principle, which safeguards the working with concepts extended so as to comprise the whole universe.

All this shows that it is by no means the idea of linguistic philosophy as another branch of critical philosophy and in Kant's sense as a prolegomenon to any metaphysical and moral philosophy which is attacked in this paper, but the neglect by linguistic philosophers of the recent development of linguistics as a science.

References

Boole, G. (1854). An Investigation of the Laws of Thought (London).

Flew, A. G. N. (1956). Essays in Conceptual Analysis (London).

Gellner, E. (1960). Words and Things (London).

Herdan, G. (1956). Language as Choice and Chance (Groningen).

Herdan, G. (1960). Type-Token Mathematics: A Textbook of Mathematical Linguistics (The Hague).

Howes, F. (1960). Letter to the Editor. *Times Lit. Suppl.*, Jan. 15th.

Urmson, J. O. (1960). Some questions concerning validity. In A. G. N. Flew, *Essays in Conceptual Analysis* (London).

Wilkinson, D. H. (1959). Towards new concepts—elementary particles. In *Turning Points of Physics* (Amsterdam).

OBJECTIVES AND TECHNIQUES OF SPEECH SYNTHESIS*

GORDON E. PETERSON AND EVA SIVERTSEN
University of Michigan

The synthesis of speech is discussed as one of the simpler problems of language automation While ultimately speech synthesizers will doubtless have many practical applications, their chief value at present is in basic research on the relation of speech parameters to linguistic judgments. Two basic methods of speech synthesis are considered : 1) the generation of speech from stored segments, and 2) the generation of speech through continuous control of the various speech parameters individually ; in the latter case, the parameters may be physiological or acoustical. It is concluded that electronic analogues of the physiological speech mechanism provide a means of evaluating hypotheses about the physiologic - acoustic speech transformation, and that acoustical speech simulators are the most realistic and practical research tools for the experimental study of speech perception.

The mechanical synthesis of speech is a subject with a long history. Some of the early attempts were mere fakes. More serious work on speech synthesis did not start till the 18th century, when several major scientists used synthesis as a tool in their research on speech production. Abbé Mical's talking heads, somewhere between 1750 and 1780, may have been the first attempt. A major advance in the development of speech synthesizers was realized in the mechanical device constructed by von Kempelen. At the same time, the Russian Kratzenstein worked along similar lines. In the 19th century, new approaches to the problem were attempted. Helmholtz worked with a series of tuning forks and resonators. Koenig tried shaped sirens, and Stumpf set up a series of organ pipes. During this century, major contributions included the organ pipes of Miller, the plasticene vowel models of Paget, and Firestone's electrical organs.

A survey of these early attempts at speech synthesis, with appropriate references, is given by Dudley and Tarnoczy (1950), and earlier by Russell (1928) ; additional information may be found in publications by Scripture (1902) and Halle (1959).

Since 1930, most serious attempts at speech synthesis have been along the lines of electrical circuits. Stewart (1922) may have been the first to attempt to synthesize speech sounds with an all-electrical network.

A more modern approach to speech synthesis was achieved by Dudley in the development of the Voder (Dudley, Riesz, and Watkins, 1939). This instrument was the first electro-acoustical device which produced sound sequences approximating the dynamics of actual speech, and was exhibited widely some years ago under the

* *This research was supported by the Information Systems Branch of the Office of Naval Research of the United States Navy, under contract Nonr 1224(22), NR 049-122.*

This paper has grown out of discussions with many colleagues. In particular, the authors would like to express their gratitude to the following for their helpful comments concerning the general rationale and instrumentation of speech synthesis : Dr. F. S. Cooper of Haskins Laboratories, New York ; Dr. C. G. M. Fant of the Royal Institute of Technology, Stockholm, Sweden ; Mr. W. Lawrence of the Signals Research and Development Establishment, Christchurch, England.

sponsorship of the Bell Telephone Laboratories.

Surveys of recent work in speech synthesis are given by Barney and Dunn (1957), Fant (1958b), and Peterson (1958).

Language automation

It is now generally recognized that automatic speech synthesis is only one aspect of the broader field of language automation. Since there are several related objectives in language automation, it seems desirable to consider their relation before considering the problem of speech synthesis in detail. Actually, speech synthesis is probably the simplest problem of language automation, primarily because it involves only the regeneration of the acoustical speech waveform. Various electronic speech synthesizers have now been constructed, and some of these have been very successful.

Vocoders

The transmission of intelligible speech over channels with low information capacity is probably second in order of basic difficulty among the problems of language automation. Systems for such transmission are commonly designated vocoders, and involve both speech analysis and speech synthesis. In the *physical parameter* systems, some type of systematic sampling of the speech signal is employed in the analysis. The samples are coded and transmitted to the synthesizer which recreates the speech at the distant end of the transmission link. Examples of physical parameter systems are the time-frequency scanning device described by Subrahmanyam and Peterson (1959) and the time-frequency compression-expansion system described by Fairbanks, Everitt, and Jaeger (1954).

In *speech parameter* systems, basic acoustical properties of the speech signal which are determined by the nature of the vocal mechanism are derived by the analyzers and are coded for transmission. The parameters include the fundamental voice frequency and the frequencies of the various formants. Examples of speech parameter vocoders are the formant-coding speech compression system of Flanagan (1956a, 1956b), and the formoder of Chang (1956) and Howard (1956).

It does not seem likely that all of the basic acoustical parameters of speech are of equal significance in all languages, and therefore at present we do not know the extent to which the same set of speech parameters will be satisfactory for different languages. Physical parameter systems, on the other hand, sample the speech of various languages independently of the significance of these parameters.

Speech parameter systems will probably provide a greater reduction in channel capacity for the transmission of intelligible speech, but routine physical sampling of the speech signal is less difficult than extraction of the basic speech parameters. Under certain conditions, some noise may resemble particular speech parameters in such a way that it cannot be distinguished from the corresponding speech parameters ; for example, a burst of noise might resemble a plosive pulse or a fricative consonant. Since the operation of a physical sampling system is independent of the particular parameter values which occur within the speech signal, it seems reasonable that such a system should be less sensitive to noise interference.

Automatic speech recognition

The automatic recognition of speech not only requires the extraction of the information-bearing parameters, but it also requires the reduction of these parameters to a discrete linguistic code. Thus automatic speech recognition involves a linguistic interpretation of the acoustical parameters of speech. Probably the essential reason that automatic speech recognition stands relatively high in the hierarchy of complexity in language automation is that it involves the conversion from continuous signals to a discrete code.

Mechanical translation

As a final area of language automation we may consider the automatic translation of languages (Bar-Hillel, 1959). Present research in this field strongly indicates that mechanical translation is most readily performed in terms of the discrete units of the languages involved, i.e. the translation is most easily accomplished with printed or other discretely coded input and output rather than with speech input and output. A complete oral—oral translator is well beyond the possibilities of current communications technology ; in such a device an automatic speech recognizer is required for the input language and an automatic speech synthesizer is required for the output language of the translator. In terms of linguistic behaviour, such a device would resemble the human engaged in language translation (Peterson, 1953). It may be noted that in automatic speech recognition a function similar to the interpretation of speech by a listener is required, whereas in automatic speech synthesis a function similar to the human production of speech is involved. The parallelism suggests that such tools of language automation may serve as important research devices for the understanding of speech reception and speech production.

Applications of speech synthesis

Until recent years it appears that applied objectives (at least beyond entertainment) have not been emphasized in the field of speech synthesis. Rather, the primary objective of research on speech synthesis has been to obtain fundamental information about the nature of speech. While important applied objectives have now been identified, it appears that the quest for basic information about speech is still the dominating motivation for such research.

Thus far the major practical objective of speech synthesis has been to convert control signals to an acoustical output in the transmission of intelligible speech over channels with low information capacity. These signals may be transmitted in either a continuous or discrete form to control the synthesizer.

There are various applications for speech synthesis in which the control signals are not derived from input speech, but are elements of a discrete code. *Telegraph* *speech* (Harris, 1953b) is an example, in which discrete control signals are transmitted to a synthesizer to generate a spoken message. In telegraph speech a human operator might be employed to initiate the message, which after transmission would operate a synthesizer automatically at the receiving terminal.

A somewhat different application would involve the *conversion of printing to speech.* An automatic system would involve several complicated operations. First the print must be read or transduced to a set of control signals. Since these signals represent conventional spellings, it is necessary to transform them to a new discrete code which represents the phonemes and prosodemes of speech. For a language such as English, a very large computer operation would be required even to approximate this transformation. Finally, the discrete code of phonemes and prosodemes must be converted to continuous acoustical signals by means of a speech synthesizer.

There are various approaches which might be followed in developing a set of rules and exceptions for converting printed English into phonemes and prosodemes. A comparative investigation of various approaches should be of considerable interest. It may be noted that grammatical and (or) semantic considerations would be necessary for completely accurate print reading. There is no other way to distinguish among homographs, e.g. *read* [rid] and [rɛd], *content* ['kantɛnt] and [kən'tɛnt]. Even greater problems are the extraction of the proper stress, intonation, and pause patterns, all of which are essential to intelligibility. With current technology it is thus unrealistic to attempt to convert conventional printing to an acoustical output which has all of the functionally distinctive properties of natural speech.

As indicated above, the synthesis of speech from a discrete code is also essential to automatic *oral-oral language translation.* There are doubtless many specific applications of speech synthesis which will eventually become common in communications technology. It is the use of speech synthesis as a research technique in the further understanding of speech parameters, however, with which this paper is primarily concerned.

Methods of speech synthesis

There are at least two general approaches to the design of a speech synthesizer. In one approach, discrete segments are connected together to produce speech. These segments may be obtained from actual human utterances or they may be discrete segments which are artificially generated by some mechanical or electro-acoustical device.

In the other general type of speech synthesizer, functions which are continuous over short intervals of time are employed to control the synthesizer. These functions may represent the basic parameters of speech production, or they may represent the various acoustical parameters of speech which result from a complicated transformation of the activities of the vocal mechanism.

Stored segments

In the first type of speech synthesizer, it is assumed that segmentation in time of the speech continuum is possible. However, such segmentation has long been a problem of speech analysis (Liberman *et al.*, 1959 ; Lehiste and Peterson, 1960). There is no simple way, on the acoustical or on the articulatory level, to segment the speech continuum into a succession of phones, syllables, or larger units. The same basic difficulties are involved in speech synthesis from discrete units.

The basic segmental unit employed may differ markedly ; the following are some of the possibilities :

Separate vowel and consonant segments. Since successive vowels and consonants have an interaction, it is not possible to synthesize normal speech with a simple set of segments in which each vowel and consonant is represented only once. Rather each vowel and each consonant phoneme must be represented in the segment catalogue several times according to the various influences of the phonetic environment in which it can occur. This general approach was investigated by Harris (1953a).

Dyads. An approach to speech synthesis was described by Peterson, Wang, and Sivertsen (1958), in which each segment extends from the steady state or target position of one phoneme to the steady state or target of the next. Thus the dyad is defined as the set of all segments involving a single articulatory target pair and all conditions of prosody associated with that pair. The individual dyads contain the transitions or speech dynamics. By associating a series of different intonations and stresses with each articulatory target pair, it is possible to include the various prosodic conditions necessary for the synthesis of normal speech. The segment inventory required for synthesizing speech according to dyads is considerably greater than that required for synthesis according to separate vowel and consonant segments, but it is considerably less than that required for some of the larger types of segments discussed below.

Half-syllables. A third approach involves the use of half-syllables, extending from the syllable initiation to some central part of the syllable nucleus or from that point to the syllable termination. Such a unit will probably include more of the speech dynamics than does the dyad, and consonant clusters can be managed more easily. On the other hand, the segment inventory will probably be larger. It will be necessary to determine the catalogue of half-syllable length articulatory sequences required to synthesize any normal utterance, and the various prosodies associated with them.

Syllables. A fourth possible choice for a basic unit in speech synthesis is the syllable. It has been assumed in the above discussion that the syllable in English can be defined. There have been various attempts to define such a unit. Respiratory (Stetson, 1951), articulatory (de Saussure, 1916), acoustical (Stowe, 1958), and distributional (Haugen, 1956) bases have been suggested. An essential problem in defining English syllables is the identification of certain syllable boundaries, for example in words with single separating consonants such as *obey* and *facet.* In these cases, for the purposes of segmentation in speech synthesis, however, the syllable divisions can be made at the consonant targets, so that half-consonants are associated with these syllables. Regardless of the conceptual basis of the syllable employed, each syllable will have to be represented by several units in the segment inventory, depending on the number of prosodic conditions which can be associated with it. The syllable is likely to contain still more of the speech dynamics than the half-syllable ; on the other hand, the inventory of required units will be still larger.

Syllable dyads. The concept of the dyad may be extended to the syllable, in such a way that the segments extend from some central point in the nucleus of one syllable to a similar point in the nucleus of the adjacent syllable. In this case, not only are syllable-initial and syllable-final consonant dynamics included, but also the dynamics of the entire intervocalic consonant sequences are subsumed. Again, it is necessary to take account of all prosodic conditions associated with each sequence, and the segment inventory will be further increased in size.

Words. The study of sound spectrograms has demonstrated that there are many intereffects across word boundaries (Lehiste, 1959). Probably the problem of segment junctures would be reduced somewhat, however, if words were employed as the basic units for synthesis. A very large segment catalogue would be required for general speech synthesis, and this would be further multiplied by the fact that many different conditions of prosody may be associated with any given word. In many applications, however, only a restricted vocabulary is required, so that the segment inventory could be considerably reduced.

Synthesis based on a segmentation of the speech continuum will in any case be built up of discrete units, and a speech synthesizer employing such units could have a discrete or quantized input, e.g. punched tape or a keyboard. Joining the segments with a minimum of disturbance is a major instrumental problem in such synthesis. More basically, in actual speech idealized target positions are rarely realized. Thus, regardless of the segmental unit (separate vowel and consonant segment, dyad, half-syllable, syllable, syllable dyad, or word) there are intereffects in normal speech which cannot be realized in the synthesis unless the catalogue is elaborated with innumerable allophonic variations. It is obvious that the simplification of employing idealized target positions will cause the speech to be less natural ; whether it will reduce or enhance intelligibility is a subject for future investigation.

While the use of increasingly larger units in speech synthesis incorporates more and more of the dynamics of speech, there is also an associated increase in the number of segments required. Further, it would seem that as the segment length is increased, less information about the details of the basic dynamics of speech will result. If information about the nature of speech is the primary objective of speech synthesis, then relatively short segments would appear to be most effective for research with stored speech segments.

Synthesis from stored segments provides one method of testing hypotheses concerning language and speech. For example, the invariance and the allophonic variations of phonemes, and the analysis of intonation patterns as combinations of a limited number of pitch levels, might profitably be studied in this manner.

Continuous parameters

A great deal of information about the allophonic variations of phonemes and prosodemes in speech may be derived from experiments with discrete segments. But this procedure is less suited for studies of the individual speech parameters. The various parameters of the speech signal must be properly represented and organized in the segments if intelligible natural speech is to be achieved. With a stored segment procedure, however, the various parameters are all contained within the time segments, and cannot be adjusted individually and continuously. For some research purposes it is more valuable to be able to adjust the individual speech parameters independently and accurately, and for this type of research a speech synthesizer which generates the parameters mechanically or electronically is required.

There are two basic types of device for such speech synthesis. This dual possibility results from the fact that there is not a one-to-one correspondence between the physiological parameters of speech production and the output acoustical speech wave. Rather, the acoustical wave represents a complicated transformation of the speech physiology. Thus, a synthesizer may employ uni-dimensional controls of either the physiological or the acoustical parameters of speech.

Speech analogues. A simulation of the actual human vocal tract may be constructed for generating speech waves. Such a device is an electro-acoustical equivalent of the physiological mechanism, and so its controls are organized in terms of the basic physiological structures or parameters of speech production.

If we take the view that the physiological mechanism is the basic information source in speech, then the essential parameters of the physiological mechanism must be considered the basic information-bearing parameters. For research purposes, output speech waves from a physiological analogue may be analyzed acoustically to determine the relationship between physiological formations and acoustical patterns. The output waves may also be presented to subjects in studies of the relation between speech production and speech perception, and of the relation of the acoustical parameters to speech perception. It is not an easy matter, however, to adjust the physiological parameters to obtain a specified sequence of acoustical parameters. Since the physiological to acoustical transformation is complicated, a physiological parameter system is limited for research on the acoustical parameters of speech. It offers the advantage, however, that the generated parameters are correct and realistic, not dependent upon the accuracy of a simplified acoustical theory of the vocal tract.

There are at least two forms which a physiological parameter system might take. The system might be constructed as a mechanical device which directly generates speech waves. Attempts to construct such a system with modern engineering techniques have not been made, and it is not easy to predict the extent of the difficulties that would be involved in the development of such a mechanism.

The speech mechanism is three-dimensional in its spatial properties, and it would obviously be very difficult to control each detail of such a three-dimensional system as a function of time. It would seem appropriate to represent each significant physiological parameter as a separate function of time. The terminology of descriptive phonetics is adequate evidence that there are a large number of such parameters, e.g. front, open, lateral, palatalized, rounded, nasal, voiced, etc. It would appear to be very difficult to control and co-ordinate these various parameters in a mechanical simulation of the human vocal tract. Whatever the exact nature of such a device might be, it would be imperative to design the controls of the mechanism in such a way that they could be easily managed for research purposes.

An alternative is the construction of an electrical analogue to the physiological system. The modern electro-acoustical approach to speech synthesis by means of vocal tract analogues was first developed in a substantial way by Dunn (1950). At the Bell Telephone Laboratories Dunn developed an electrical model of the vocal tract which could simulate variable positions of the tongue constriction and variable lip rounding for producing vowel sounds. His accompanying theory of vowel production, including a mathematical explanation of the basis of the various formant frequencies, was a major advance in the field of speech communication. Soon thereafter, Stevens, Kasowski, and Fant (1953), working at the Massachusetts Institute of Technology, developed a more complete analogue, including variable cross-sections of the vocal tract, and further refined the theory of vowel formation to include dissipation in the vocal cavities.

Since the time of these original developments, several others have made important contributions to the concept of analogue speech synthesis. These include the work of House and Stevens (1955) and Weibel (1955). Fant (1958a) has shown in detail the calculations upon which such analogues can be built, and developed further his own analogue LEA. Rosen (1958) has made an important contribution to the development of dynamic, as opposed to static, speech analogues.

Thus far the analogues which have been constructed are partial analogues. A complete electrical analogue of the human vocal system would be a very complicated device. In the complete device it would be necessary to simulate the thorax and the larynx as well as the supraglottal cavities and articulators. Difficult problems include the generation of plosive and fricative consonant sounds. In a complete analogue these sounds should not be produced by isolated impulse and random noise generators, but should be the result of adjustments of the equivalents of the speech organs according to the manner of consonant sound production in actual speech. The overall topology of such a device seems no more complicated than the dynamic considerations which would be essential to its proper function. In other words, since speech involves a great deal of movement and variation, it would be important to obtain a reasonably natural regulation of the dynamics of the system.

The great complexity involved in the complete simulation of the human vocal tract makes it impractical with current technological methods to construct a complete physiological analogue. There is not sufficient information about the operation of the human vocal tract to insure the accuracy of such an analogue. Thus, it appears that information about the transformation from physiological to acoustical parameters cannot be derived most readily from such a synthesizer. Rather, such questions should be answered by modern techniques of research applied directly to the human vocal mechanism. The use of electromyography, the measurement of subglottal pressures, the use of pressure and velocity probes, and direct and x-ray motion pictures are all techniques which may be of aid in specifying the detailed activities of the human vocal mechanism. Thus it would appear that research on the transformation from speech physiology to speech acoustics is primarily a subject for instrumental analysis rather than for speech synthesis. The construction of restricted analogues, however, provides an excellent means of testing and evaluating research findings about the activities of the human vocal mechanism in speech production.

Acoustical speech simulators. If properly designed, a speech synthesizer may provide the basis for important information about the perception of the various acoustical parameters of speech. It is possible to simulate the acoustical speech signal in various ways. Acoustical speech simulators have no simple correspondence to the physiological vocal tract, although they provide an approximation to some desired set of properties of the acoustical speech output. For research purposes it seems that the most useful technique will involve a simulation of the various basic acoustical speech parameters. It should be emphasized that since acoustical simulators are one step more removed from the physiological source, it is only with extreme caution, if at all, that one can draw conclusions about the behaviour of the physiological mechanism from studies with such simulators.

Various types of acoustical speech simulators have been developed. The " pattern playback ", designed by Cooper (1950) at the Haskins Laboratories, may be considered such a speech simulator. This instrument is not an electrical analogue of the vocal tract, but is a tone generator so designed that the amplitudes of the various harmonics can be controlled to simulate the acoustical waves of speech. Among speech synthesizers, this is the instrument which thus far has been used most extensively for research on speech. The essential method of these studies is to obtain listener classifications of various acoustical waves into linguistic categories. A recent summary of experimental results obtained by this method is given by Liberman (1957).

Lawrence (1953) was the first to develop a dynamic speech synthesizer employing a basic set of essentially independent acoustical speech parameters which are closely related to those generated by the human vocal tract. The original model developed by Lawrence employed parallel resonators to simulate the formant frequencies, but various discussions since have suggested that a series (cascade) type of resonator design may more closely simulate the acoustical speech wave. The basic technical considerations have been outlined by Flanagan (1957). Also of considerable interest is the work of Fant (1959), who has developed simulators which may be either manually (OVE I) or electronically (OVE II) controlled.

Another type of device, which is remotely related to the technique of dyad synthesis discussed earlier in the present paper, has been developed both by the Haskins Laboratories (Borst, 1956) and by Fry and Denes (not yet described in publication). In these devices, sequences of steady-state values may be specified for a number of different acoustical parameters, so that they might be called acoustical target simulators. The transitions between the targets are electronically controlled, and in general any type of smoothing function between successive targets may be employed.

Evaluation of speech synthesizers

The problem of evaluating speech synthesizers is not a simple matter. One of the difficulties is simply that with the instruments which have thus far been constructed, it is very difficult to synthesize large amounts of material for evaluative tests.

It has long been demonstrated that monosyllables provides the most exacting test of speech intelligibility (Fletcher and Steinberg, 1929). Certainly, one crucial test of a speech synthesizer would be to compare its intelligibility in generating monosyllables with that which can be achieved with natural speech. Perhaps under some circumstances even more crucial tests can be obtained by comparing the intelligibility in the two cases in a specified background of random noise.

Various research workers have occasionally been disturbed over the fact that monosyllabic intelligibility tests do not contain complicated prosodic information. Under prosodic information the authors would like to include quantity (duration), stress, tone, intonation, and vocal quality. These parameters are not consistently evaluated in tests with monosyllables, but if desired they could be systematically incorporated into such tests to a much greater extent than they have been in the past. It is our opinion that, in general, intelligibility tests in a background of noise will be as discriminating among various types of synthetic speech as will judgment tests of naturalness. Such intelligibility tests may provide critical evaluations, but they are not diagnostic in nature. Thus various types of comparative judgment tests between synthesized utterances and normal utterances may provide a more effective technique for determining specific faults in speech synthesizers. Many different psycho-physical methods might be employed in such tests. Probably the chief reason that such comparative tests have not yet been employed systematically is that most synthesizers have not yet been able to produce speech which sufficiently approximates the normal to make the tests particularly informative.

Conclusion

Two general types of speech synthesis have been considered: 1) the generation of speech from stored segments, and 2) the generation of speech from continuous parameters. A number of different types of segments, varying considerably in length, have been considered for the synthesis of speech. It appears that most information about the nature of speech is to be derived from the use of relatively short segments.

Throughout this paper it has been emphasized that the acoustical parameters of speech represent a complex transformation of the physiological parameters. For research purposes, it would seem that an effective physiological analogue to the human speech mechanism should employ unidimensional input controls which specify the various basic physiological parameters involved in speech production. Only a physiological simulator is an actual analogue of the human vocal tract. Complete electronic analogues of the human speech mechanism have not yet been constructed. Thus far, the analogues have been of a restricted nature, actually representing only certain aspects or subdivisions of the complete vocal system. The chief research value of such analogues of the physiological mechanism may be to check experimental information about the structure and function of the vocal mechanism.

It has been suggested that electro-acoustical simulators of the speech signal are most useful for studying the linguistic categories into which listeners classify speech waves containing various parameter values. Obviously, an effective acoustical simulator should employ uni-dimensional input controls to represent the various acoustical parameters of speech.

References

Bar-Hillel, Y. (1959). Report on the State of Machine Translation in the United States and Great Britain (Unpublished technical report. Jerusalem).

Barney, H. L. and Dunn, H. K. (1957). Speech synthesis. *Manual of Phonetics* (Amsterdam), 202.

Borst, J. (1956). The use of spectrograms for speech analysis and synthesis. *Journal of the Audio Engineering Society*, **4**, 14.

Chang, S-H. (1956). Two schemes of speech compression system. *J. acoust. Soc. Amer.*, **28**, 565.

Cooper, F. S. (1950). Spectrum analysis. *J. acoust. Soc. Amer.*, **22**, 761.

De Saussure, F. (1916). Cours de linguistique générale (Lausanne, Paris).

Dudley, H. W., Riesz, R. R. and Watkins, S. S. A. (1939). A synthetic speaker. *Journal of the Franklin Institute*, **227**, 739.

Dunn, H. K. (1950). The calculation of vowel resonances and an electrical vocal tract. *J. acoust. Soc. Amer.*, **22**, 740.

Fairbanks, G., Everitt, W. L. and Jaeger, R. P. (1954). Method for time or frequency compression-expansion of speech. *Transactions of the IRE-PGA*, AU2, 7.

Fant, C. G. M. (1958a). Acoustic theory of speech production. Royal Institute of Technology, Division of Telegraphy - Telephony, Report No. 10 (Stockholm).

Fant, C. G. M. (1958b). Modern instruments and methods for acoustic studies of speech. *Proceedings of the VIII International Congress of Linguists* (Oslo), 282.

Fant, C. G. M. (1959). Acoustic analysis and synthesis of speech with application to Swedish. *Ericsson Technics*, 1.

Flanagan, J. L. (1956a). Automatic extraction of formant frequencies from continuous speech. *J. acoust. Soc. Amer.*, **28**, 110.

Flanagan, J. L. (1956b). Development and testing of a formant-coding speech compression system. *J. acoust. Soc. Amer.*, **28**, 1099.

Flanagan, J. L. (1957). Note on the design of " terminal-analog " speech synthesizers. *J. acoust. Soc. Amer.*, **29**, 306.

Fletcher, H. and Steinberg, J. C. (1929). Articulation testing methods. *Bell System Technical Journal*, **8**, 806.

Halle, M. (1959). The Sound Pattern of Russian ('s-Gravenhage).

Harris, C. M. (1953a). A study of the building blocks of speech. *J. acoust. Soc. Amer.*, **25**, 962.

Harris, C. M. (1953b). A speech synthesizer. *J. acoust. Soc. Amer.*, **25**, 970.

Haugen, E. (1956). The syllable in linguistic description. *For Roman Jakobson* (The Hague), 213.

House, A. S. and Stevens, K. N. (1955). Auditory testing of a simplified description of vowel articulation. *J. acoust. Soc. Amer.*, **27**, 882.

HOWARD, C. R. (1956). Speech analysis-synthesis scheme using continuous parameters. *J. acoust. Soc. Amer.*, **28**, 1091.

LAWRENCE, W. (1953). The synthesis of speech from signals which have a low information rate. *Communication Theory* (London), 460.

LEHISTE, I. (1959). An Acoustic-Phonetic Study of Internal Open Juncture. The University of Michigan Speech Research Laboratory Report Number 2 (Ann Arbor).

LEHISTE, I. and PETERSON, G. E. (1960). Duration of syllable nuclei in English. To appear in *J. acoust. Soc. Amer.*, **32**.

LIBERMAN, A. M. (1957). Some results of research on speech perception. *J. acoust. Soc. Amer.*, **29**, 117.

LIBERMAN, A. M., *et al.* (1959). Minimal rules for synthesizing speech. *J. acoust. Soc. Amer.*, **31**, 1490.

PETERSON, G. E. (1953). Basic physical systems for communication between two individuals. *J. Sp. and Hear. Dis.*, **18**, 116.

PETERSON, G. E. (1958). Fundamental problems in speech analysis and synthesis. *Proceedings of the VIII International Congress of Linguists* (Oslo), 267.

PETERSON, G. E., WANG, W. S-Y. and SIVERTSEN, E. (1958). Segmentation techniques in speech synthesis. *J. acoust. Soc. Amer.*, **30**, 739.

ROSEN, G. (1958). Dynamic analog speech synthesizer. *J. acoust. Soc. Amer.*, **30**, 201.

RUSSELL, G. O. (1928). The Vowel (Columbus, Ohio).

SCRIPTURE, E. W. (1902). The Elements of Experimental Phonetics (New York).

STETSON, R. H. (1951). Motor Phonetics. 2nd ed. (Amsterdam).

STEVENS, K. N., KASOWSKI, S. and FANT, C. G. M. (1953). An electrical analog of the vocal tract. *J. acoust. Soc. Amer.*, **25**, 734.

STEWART, J. Q. (1922). An electrical analogue of the vocal organs. *Nature*, **110**, 311.

STOWE, A. N. (1958). The Syllable in Linguistics and Automatic Speech Recognition (Unpublished dissertation, Harvard University, Cambridge, Massachusetts).

SUBRAHMANYAM, D. L. and PETERSON, G. E. (1959). Time-frequency scanning in narrow band speech transmission. *IRE Transactions on Audio*, AU-7, 148.

WEIBEL, E. S. (1955). Vowel synthesis by means of resonance circuits. *J. acoust. Soc. Amer.*, **27**, 858.

EFFECTS OF DELAYED VISUAL CONTROL ON WRITING, DRAWING AND TRACING

H. Kalmus, D. B. Fry and P. Denes
University College, London

Writing, drawing to verbal instruction and tracing were recorded on a combination of a telescriber with a short-term information store resulting in visual delay. The resulting conflict of visual and kinaesthetic feed-back slowed down performance and resulted in errors such as overshooting, repetition and wrong spacing. Measurable effects such as the duration of writing and the error area in tracing could be shown over the observed range to increase with the amount of delay.

These results are compared with those of previous experiments in the visual and acoustic fields and certain similarities and differences are discussed.

Introduction

The first experiments employing delayed sensory feed-back were performed in the acoustic field (Lee, 1950, Black, 1951, Fairbanks and Guttman, 1958). In these experiments speech produced by a person reading aloud is replayed to him through earphones after a variable delay. With delays between 0·1 and 0·6 sec. the interference with normal speech is considerable. The subject becomes uncertain and speaks haltingly; he will, under such conditions, make articulatory errors or slow down in his performance, or both. Delay in acoustic feed-back also affects non-vocal activities such as clapping one's hands, playing a musical instrument (Kalmus, Denes and Fry, 1955) or finger tapping (Chase *et al.*, 1959).

It was thought that comparable effects might be produced by visual delay and that the most promising activity for investigation might be writing, which is analogous to speech. This idea had, unknown to us, occurred also to van Bergeijk and David (1959) at the Bell Telephone Laboratories. Their apparatus consisted of a telewriter, the stylus of which was made to produce a delayed trace on a Hughes Memoscope which was watched by the writing subject. Photographs of the resulting oscillograms were judged for " neatness " by a panel of judges and certain broad conclusions were based on the scores produced with different amounts of delay and under varying instructions. This work will be discussed together with our more detailed results at the end of this paper.

The functional structure of writing

Writing may be considered as an organised or patterned activity which is centrally released and peripherally controlled. Writing can furthermore be described as a sequential production of letters and other signs grouped into words and sentences. The sequence of letters in a word is in a literate person fairly predictable and under voluntary control, but the size, shape and spacing of the letters which are the result of writing are not so easily controlled and are to some extent characteristic for an individual ; this fact provides the basis for graphology which is an attempt to interpret intuitively characteristics of an individual's writing in terms of traits in his personality.

The present study of writing starts like graphology from a consideration of the shape and spacing of written letters but it is part of an attempt to interpret these characteristics generally in terms of separate sensory-motor control mechanisms.

Fig. 1 is a hypothetical scheme or functional structure (Mittelstaedt, 1954) showing the mechanisms which may be concerned with the writing activity at the relevant level.

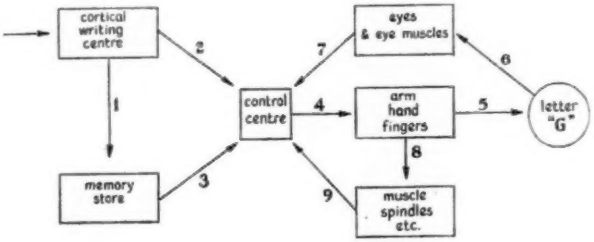


Fig. 1. Hypothetical scheme of some factors involved in the control of writing.

Interference with some of these functions (1, 2, 3) may be occasioned by various psychiatric conditions, while others may be implicated in neurological disturbances. Several of these functions may also be impaired in patients suffering from gross mutilations. Experimentally one can interfere with the visual feed-back by the application of inverting prisms or by mirrors ; here we are concerned with the effects of delaying the visual information flowing back to a writer from a letter (the letter " G ") while it is being written. This information is carried along (6) and (7) and for the time being we do not attempt to decide whether the resulting visual localisation is purely retinal or whether it is partly kinaesthetic, that is derived from the eye muscles while the eyes follow the writing trace. In any case this delayed information conflicts with kinaesthetic information originating from the writer's arm, hand and fingers by way of (8) and (9). Crudely speaking, one might say that when writing is delayed proprioception from the hand indicates that the drawing of a feature has been completed whilst the eyes and eye muscles contradict this.

This conflict can be resolved by closing one's eyes or by " not looking " and thus eliminating the visual and optomotor feed-back ; the writing movements are then solely controlled from the arm, hand and fingers and apart from accurate placing such writing appears perfectly normal (see Fig. 2). Total blocking of the visual pathway has thus only slight effects, while, as we shall see, a delay of the visual information produces much more marked writing disturbances. In drawing shapes less familiar than letters or numerals, however, visual control appears essential and it is of course impossible to place a stylus accurately or to trace anything without looking.

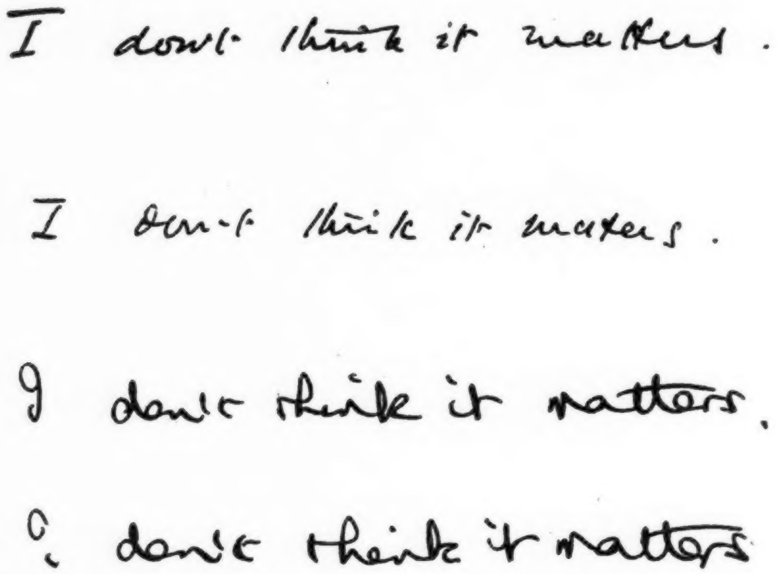


Fig. 2. Writing of two subjects with open (top lines) and closed eyes.

Apparatus

Writing was delayed by inserting a short-term information store into the control circuits of a telescriber instrument.

The telescriber, manufactured by the Telautograph Corporation, consists of a stylus which is moved by the subject and a writing pen which is linked electro-mechanically with the stylus in such a way that the pen automatically follows any movement imposed on the stylus.

The stylus is joined mechanically to the sliders of two potentiometers in such a way that one will move proportionally to a component of stylus movement in one direction while the other will move proportionally to the stylus movement in the perpendicular direction. In this way two voltages are available, one from each potentiometer, which are proportional to the components of stylus movement in two directions at right angles to each other, the two voltages defining the position of the stylus in terms of a pair of rectangular co-ordinates.

The writing pen is linked mechanically to two coils, placed in the fields of separate permanent magnets, where each coil can slide along a path at right angles to that of the other one. The voltages defining the movement of the stylus are connected one to each coil. The magnetic inter-action between coil-currents and permanent magnets will move the coils, and therefore the pen, to a position proportional to the voltages defining the stylus position and hence the pen will follow the movement of the stylus. The pen writes in ink on paper placed underneath it.

Delayed writing is obtained by putting a delay element into the link between stylus output voltage and pen-coil. The delay element consists of a number of condensers mounted on a rotating drum ; one side of all the condensers is joined together and the other sides are connected separately to segments of a commutator which in turn is in contact with two brushes. Two similar sets of condensers, commutators and two brushes are provided, mounted on the same drum, one for each of the two voltages derived from the stylus. Fig. 3 shows the arrangement for one of them.

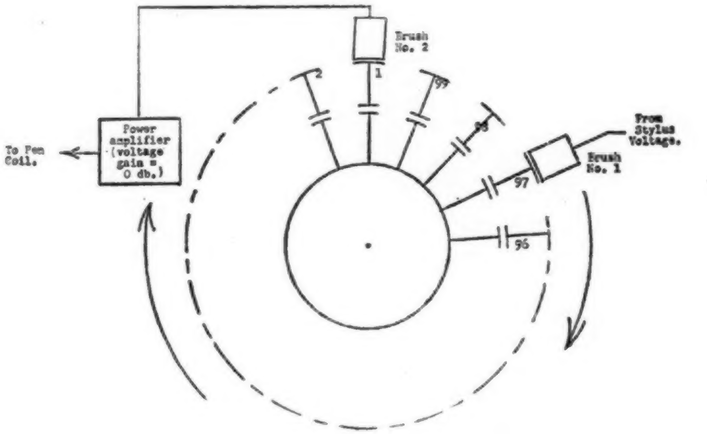


Fig. 3. Schematic representation of the delay mechanism in the apparatus described.

The varying stylus voltage to be delayed is connected to the commutator through the first brush and the condensers are charged to this voltage as they rotate past the brush. As the drum rotates, condensers 1 to 97 are charged to voltage values that represent the past history of variation of the stylus voltage connected to the first brush. The voltage on the pen-coil is determined by whichever condenser happens to be in contact with the second brush. The power amplifier shown in Fig. 3 ensures that the current for driving the pen-coil is available without discharging the condensers. In this way, the voltage on the coil follows the variations of the stylus voltage, but is delayed by the length of time it takes the condenser to rotate from the first to the second brush. The amount of delay can be changed by altering the rotational speed of the drum ; this speed is variable over a 20 to 1 range by controlling the setting on a torque converter which couples the drum to a constant speed driving motor. The brushes are mounted only 18° apart and therefore the condensers must make almost a full revolution to travel from brush 1 to brush 2. The maximum speed of the drum is about 12 revolutions per second, giving a minimum delay of about 80 msecs.; and the maximum delay is about 1·5 sec. The rotation of the drum can be reversed so that the condensers move through only 18° to get from one brush to the other and then the delay varies from 4 to 80 msecs.

A basic limitation of such a system is provided by the integrating action of the condensers. Each condenser and its associated commutator segment remains in contact with the brush for a finite length of time and integrates the stylus variations over this time-span. The design must ensure that this time quantum is short compared with the rate at which significant changes of stylus position are produced by the subject. Since there are 100 segments on each commutator, the time quantum is 0·01 of the total delay time. The effect is most serious when the delay is long ; for a delay of 1 sec., the unit of quantisation is 10 msec.; for a delay of 100 msec. the time quantum is only 1 msec. Trials have shown that this effect did not seriously limit the usefulness of the device. The effect of other time factors such as the charge and discharge time constants of the condensers and the frequency response of the circuits was made negligible by suitable electronic design.

Results

Writing experiments

Writing was done from memory or from dictation. In the first instance there was no constraint concerning time and the subjects reacted to the delay either by writing more clumsily or, in an effort to avoid this, by writing more slowly. Sometimes they combined both reactions or alternated between them (see Figs. 4 and 5). Among the most conspicuous defects of the writing were the wrong placing of words and the repetition of strokes, for instance the substitution of an " m " for an " n ". This is analogous to the repetition observed in speech (Fairbanks and Guttman, 1958) and other motor activities (Kalmus, Denes and Fry, 1955) in conditions of acoustic delay. Omissions and other non-repetitive additions were also observed. Distortion and slowing down increased more or less proportionally with the amount of visual delay over the entire range measured. This differs from the effects of acoustic delay on

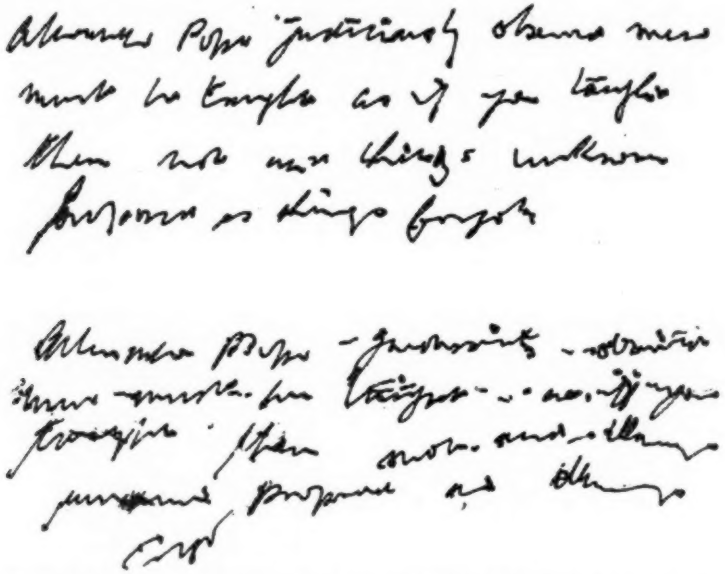


Fig. 4. Sentence written from memory in 51 sec. without delay (top) and in 57 sec. with a visual delay of 150 msec.

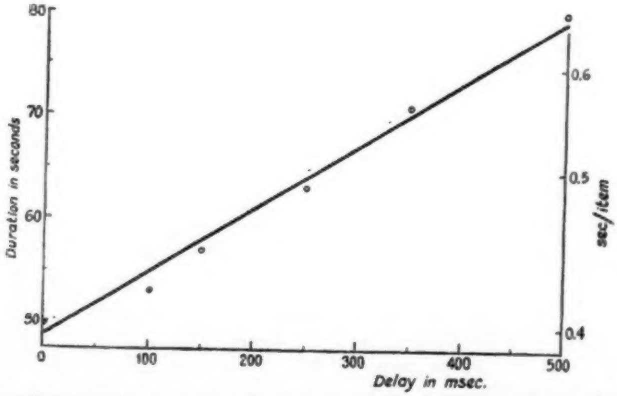


Fig. 5. Time taken to write a sentence from memory as a function of visual delay. The first two points to the left of the graph refer to the two writing samples shown in Fig. 4. For the purpose of calculating the sec./item figure, each letter and each space between words was counted as an item.

speech, which reach a maximum near the duration of a syllable, between 200 and 400 msec.; and decrease with greater delays.

In writing from dictation, speech was supplied at a controlled rate from a tape recorder. In this situation subjects could not escape into slower writing but had either to write more clumsily or to leave out bits. Otherwise there was no great difference between free writing and writing from dictation.

Drawing to instructions

The results of these writing experiments broadly confirm some of van Bergeijk and David's (1959) conclusions, but it was felt that a more detailed analysis of the effects of visual delay on handwriting would depend on further preliminary exploration. One of the methods adopted was to ask subjects to perform tasks less stereotyped than ordinary writing. Fig. 6 shows for instance the effect of visual delay on the writing of two less familiar symbols, a treble clef and a double integral ; both are grossly distorted.

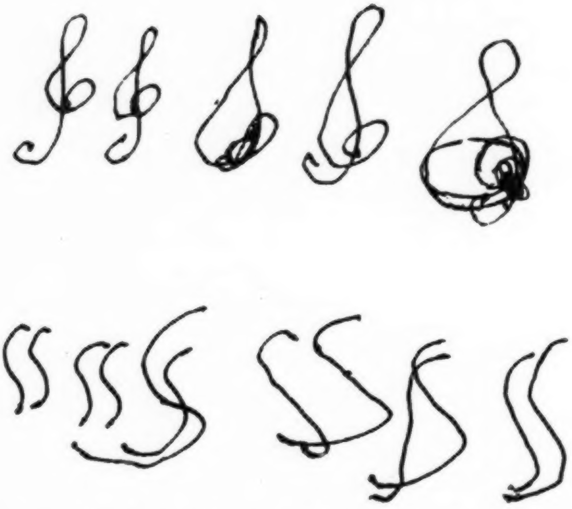


Fig. 6. Effect of visual delay on the drawing of a treble clef or a double integral sign. The first two items in each line were produced without visual delay and the remainder with a 250 msec. delay.

Inspection of fairly simple geometrical figures such as straight lines, circles or semicircles drawn while visual information was delayed, revealed a particular feature of delayed writing, namely *overshooting* (see Fig. 7), that is the continuation of a stroke beyond its intended end point after it had in fact—but not visually—been completed. Overshooting is the analogue of non-repetitive addition under acoustic feed-back as described by Fairbanks and Guttman (1958).

The drawing, in conditions of visual delay, of running polygonal graphs to acoustic

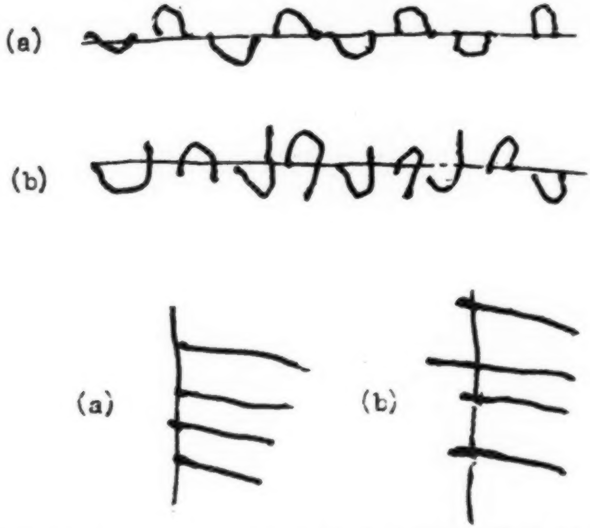


Fig. 7. Overshooting as a consequence of visual delay. Subjects were asked to draw semicircles (upper part of figure) and horizontal lines (lower part) to finish at straight lines already drawn, (a) with no delay, (b) with 250 msec. delay.

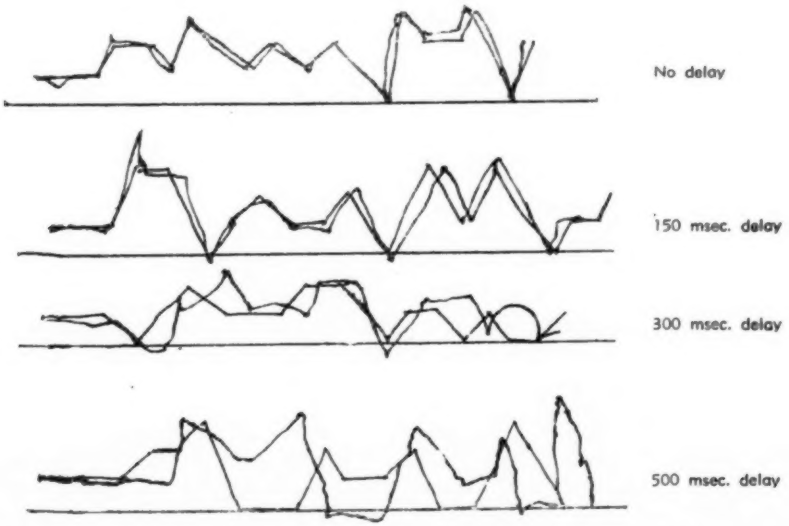


Fig. 8. Polygonal curves drawn during 25 sec. on squared paper (not shown) to tape-recorded acoustic commands. Two lines are shown for each delay: one drawn with the visual delay indicated, the other without visual delay and without a time limit.

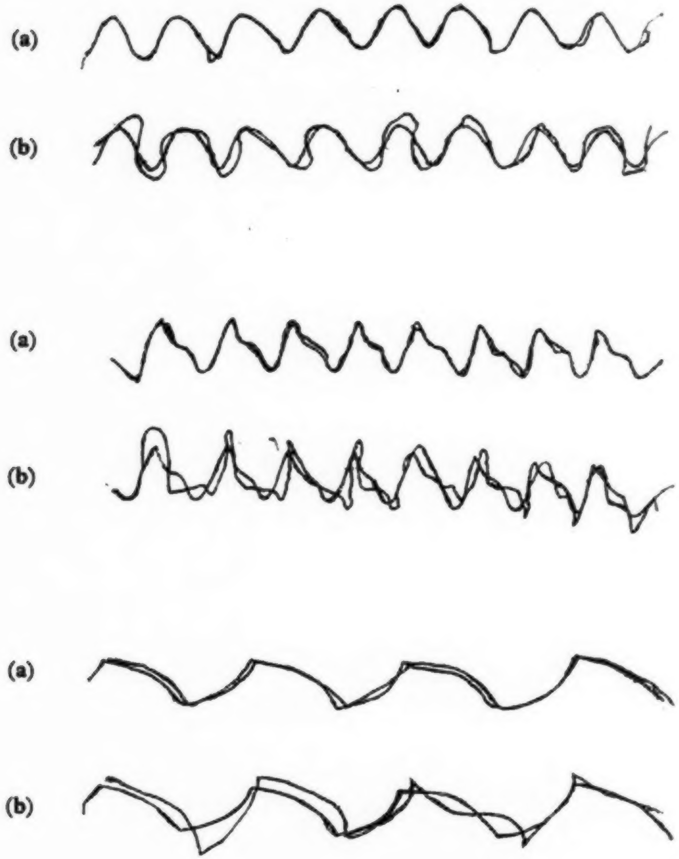


Fig. 9. Tracings of three periodic curves (a) without and (b) with visual delay of 400 msec.

command also resulted in overshooting. In addition other relational mistakes (see Fig. 8) occurred, which can be attributed to mistaken stylus positioning and to hastiness in trying to compensate for a pause. Most of these mistakes while easy to demonstrate are difficult to measure.

Tracing

Measurable effects of visual delay were produced in tracing experiments. Prepared writing models or repetitive graphs were shown on the Telautograph screen and the subjects were asked to trace them while under the influence of varied visual delay. This arrangement, by forcing the subject to look simultaneously at the original and at his own delayed copy, makes it impossible for him to " switch off " visual control for

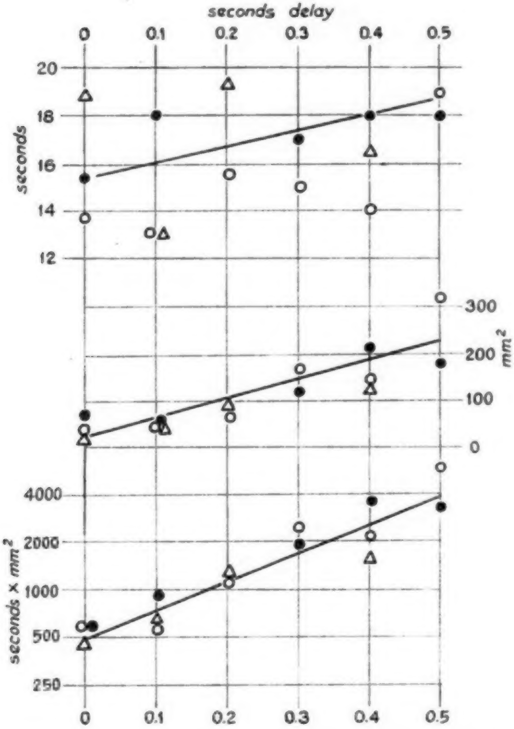


Fig. 10. Effects of visual delay on the duration and accuracy of tracings, and a combined measure of both on a logarithmic scale. The tracings were made by three male adults.

any length of time. Fig. 9 shows some results from tracing graphs. The top pair represent simple sinusoidal curves drawn with zero and with 400 msec. delay, while the two lower pairs are tracings of more complex periodic functions. It appears that the visual delay impairs accuracy somewhat in the simple curve and more so in the complex graphs. By comparing the original with the traced contours on such graphs, certain effects can be intuitively inferred, mainly overshooting, misplacing, lag, overcorrection, etc. In addition two parameters can be measured: (1) duration of performance, (2) accuracy as measured by the area enclosed between the original graph and its tracing. Fig. 10 shows with respect to one of the tracings the effects of visual delay on duration and accuracy as well as on a combination of both, namely their product. All three values increase on the average with the length of the delay ; duration and delay are only very imperfectly correlated (r = 0·338 ± 0·237), area and delay are more closely related (r = 0·837 ± 0·080) and the measure combining duration with area shows the highest correlation with delay (r = 0·886 ± 0·057).

This agrees with the observation that the time a person takes for a tracing is largely a matter for him to decide and thus greatly varies between people and on different occasions, while the inaccuracy as measured by the area is less easily controlled and is not greatly improved at the expense of the time needed for the test. In spite of these positive correlations, individual performance sometimes improved with an increase in delay and it might be thought that practice played some role in this. Writing on the machine had however been practised by the subjects for some while before the tests to which these measurements refer and no consistent improvements were noticed in repeated tests, within the repetitions of one run ; thus it rather appears that people try to vary their response, sometimes improving their performance but on other occasions making it worse.

Discussion

Qualitatively our observations of writing under delayed visual control confirm those of van Bergeijk and David (1959) in that omissions, duplications and substitutions occurred ; in addition we observed the wrong placing and spacing of words. Like them we also found that the time needed for writing increased with an increase in visual delay.

Van Bergeijk and David, in their experiments, dealt with free writing only and " neatness " was assessed by scoring and rating. In the present experiments we have introduced tracing and drawing at command so as to offer a basis for comparing the writer's intentions with his performance. Thus we have been able to measure a variable other than time—the area enclosed between the " correct ", original or intended graph and the graphs drawn under different visual delays. Duration of writing and " error area " increase with the amount of visual delay and—though the subjects manifestly varied their attempts to overcome the difficulties of the delay—we believe that the " trading " of speed for error area or vice versa was in fact not very successful. Consequently the error area is by itself a good measure of the effects of visual delay and the product of error area and time only very slightly better.

In comparing the effects of visual delay with those of auditory delay, certain similarities and also some dissimilarities become apparent. Both speech and writing are subject to dual feedback control, in the case of speech, kinaesthetic and auditory, in writing, kinaesthetic and visual. A major difference between speaking and writing is that *peripherally* the influence of the ear in the control of speech is inescapable for the hearing subject while the influence of the eye in the control of writing is not. We can close our eyes and still write fairly well but we cannot close our ears in order to evade acoustic interference. In experiments with delayed feedback, it is clear that delay will have its effect only in conditions where the subject cannot escape from the influence of the sensory feedback and rely solely on kinaesthetic feedback.

Visual delay had an effect in our experiments only when the subject was told to watch the trace (in free writing) or in tasks like tracing where he could not avoid doing so. So long as this condition obtained, then as one would expect, the longer the delay, the greater its effect on performance. In the results obtained by van Bergeijk and David, the effect of delay was still rising at 525 msec., the longest delay used in their experiments and our own results are similar. Further, in the present experiments, at delays between 500 msec. and 1 sec. the subjects' performance became so poor as to make measurements impossible but it was evident that no improvement was to be expected within this range.

In the case of delayed auditory feedback, the effect of the delay on speech is maximal at delays between about 200 and 400 msec. Subjects who undergo delayed auditory feedback usually report that, when the delay is long enough, they are able to dissociate the acoustic signal from their own speech activity ; the experience becomes akin to that of speaking in a highly reverberant room. In this case, the ear is still receiving the acoustic stimuli that are fed back (as we have seen, it cannot escape them) but the brain suppresses the connection between them and the motor activity of speech. There are some subjects, in fact, who can effect the suppression even with short delays, though this requires a great effort. The improvement in speech performance noted with long delays is due, therefore, to the fact that the subject now escapes from the effect of auditory delay by central inhibition. He controls his speech mainly through kinaesthetic feedback and his position is analogous to that of the subject writing with his eyes closed.

If we compare visual delay in writing and auditory delay in speech in conditions where each is working, then the effects are very similar. Both produce a conflict with kinaesthetic feedback and as a result both activities are slowed down and analogous mistakes such as repetitions, other additions and omissions appear. The choice of syllables in speech and of letters in writing are under central control and are little affected by feedback whereas the details of their execution are dually monitored and are thus affected by the conflict between the immediate kinaesthetic information and the delayed acoustic or visual information.

We wish to thank the University of London for a grant from the Central Research Fund and the Telautograph Corporation of Los Angeles for a considerable reduction in the cost of the instrument used.

REFERENCES

BERGEIJK, W. A. VAN and DAVID, E. E., JR. (1959). Delayed handwriting. *Perceptual and Motor Skills*, **9**, 347.

BLACK, J. W. (1951). The effect of delayed side-tone upon vocal rate and intensity. *J. Speech & Hearing Dis.*, **16**, 56.

CHASE, R. A. *et al.* (1959). A comparison of the effects of delayed auditory feedback on speech and key tapping. *Science*, **129**, 903.

DEUTSCH, J. A. and CLARKSON, J. K. (1959). Nature of the vibrato and the control loop in singing. *Nature*, **183**, 167.

FAIRBANKS, G. and GUTTMAN, N. (1958). Effects of delayed auditory feedback upon articulation. *J. Speech & Hearing Res.*, **1**, 12.

KALMUS, H., DENES, P. and FRY, D. B. (1955). Effect of delayed acoustic feedback on some non-vocal activities. *Nature*, **175**, 1078.

LEE, B. S. (1950). Effects of delayed speech feedback. *J. acoust. Soc. Amer.*, **22**, 824.

MITTELSTAEDT, H. (1954). Regelung in der Biologie. *Regelungstechnik*, **2**, 177.

MITTELSTAEDT, H. (1954). Regelung und Steuerung bei der Orientierung der Lebewesen. *Regelungstechnik*, **2**, 226.

ON THE PERCEPTION OF UNRELEASED VOICELESS PLOSIVES IN ENGLISH

BJÖRN STÅLHANE ANDRÉSEN
University of Bergen

An experiment concerned with the possibility of distinguishing auditively between unreleased /ʔ/ and unreleased /p/, /t/, and /k/ in English was carried out, with two groups of hearers, one listening by means of earphones, one listening by means of loudspeaker. It was found that correct identification of all the voiceless stops largely depended on the way in which the sounds were (re)produced and conveyed to the hearers. It was also found that correct identification very much depended on the quality of the preceding vowel sound, especially in the case of /p/, /k/, and /ʔ/. Under the best listening conditions /ʔ/ was identified with considerable accuracy. Under less favourable conditions the identification was—in part—scarcely better than pure chance.

INTRODUCTION

In the autumn of 1955 Fred W. Householder, Jr. made an experiment (the last of a series made by him) the aim of which was to determine to what extent final unreleased (or inaudibly released) /p/, /t/, and /k/ can be distinguished in American English (Householder, 1956). His hearers were asked to identify the final consonants of a large number of words as /p/, /t/, or /k/. Householder also included some words ending in the glottal stop, without informing his hearers of this, to see how they would respond to this sound (ibid., p. 236).

In order to determine how far it is possible for an English-speaking person to distinguish auditively between unreleased glottal stop and unreleased /p/, /t/, and /k/ respectively (and between the latter sounds mutually when the glottal stop is included in the system), the author made an experiment, to some extent along the same lines as Householder's experiment, but with more restricted material and with the principal difference that the glottal stop was included as one of the sounds the hearers were asked to identify.

The experiment was carried out at University College London in the autumn of 1959.

HEARERS AND TEXTS

The texts used for the experiment were read by Mr. A. C. Gimson of University College London, recorded on tape (by means of a *Tandberg Bandopptaker 2,* frequency range 30 - 8000 c/s, speed 3¾" per second), and played back to the hearers.

One group of hearers listened by means of earphones (*Safety Supply Co.,* Canada, and *Headset H-70/AIC,* USA). In the following these hearers will be referred to as the *Earphone Group*.

Two groups of hearers listened by means of loudspeakers. In the following these two groups will be treated together and referred to as the *Loudspeaker Group*.

Each hearer listened to each item of the material only once.

The texts consisted of lists of separate *words* ending in unreleased voiceless plosives. The texts were divided into two *sections*. The first section, in the following referred to as the 1ɔ-section, consisted of four *series* and was built up in this way :—

Series 1: The words /lɔp/, /lɔt/, /lɔk/, and /lɔʔ/, altogether 30 items, mixed at random, /lɔp, /lɔt/, and /lɔk/ occurring 7 times each, /lɔʔ/ 9 times. The hearers were told that the items would end in /p/, /t/, /k/, and /ʔ/, but were ignorant of how many there were of each type. They were asked to write down immediately the sound they thought they heard in each case.

Series 2: The words /lɔp/ and /ɔʔ/ mixed at random, altogether 24 items, each type occurring 12 times. The hearers were asked to write either *p* or ʔ.

Series 3 and 4 were built up like Series 2, except that the words to be distinguished between were /lɔt/ and /lɔʔ/, /lɔk/ and /lɔʔ/ respectively.

The second section, in the following referred to as the li-section, was constructed on the same pattern as the lɔ-section, except that the words were /lip/, /lit/, /lik/, and /liʔ/. The number of occurrences of each plosive within each series was the same in the two sections, but their *order* was different.

It will be noticed that the words used in the experiment are all English words, but they are words that are not often heard in isolation.

In the following the term *item* will be used to denote a word as it occurs as a number on the list. Thus, in Series 1 (of each section) there are 30 items. There are 7 items ending in /p/, 7 ending in /t/, and so forth. The term *signal* will be used to denote the sound waves of one item when it is spoken or reproduced, reaching the ears of one hearer. Thus, the 7 /p/-items when spoken to 14 hearers produce 98 /p/-signals, and so forth. The term *response* will be used to denote one hearer's visible reaction to one signal, apparent in his writing *p, t, k,* or ʔ. When a response is characterized as *correct,* it means *in conformity with the speaker's intention.*

The Earphone Group numbered 14 hearers. They were:—

7 postgraduate students of phonetics, two of whom were Welsh, one American, the rest English.

3 undergraduate students of English, all English.

4 people without university education, all English. (These 4 did not come to the College for the experiment, but listened to the recordings in their homes.)

There was no marked difference in the number of correct responses given by each of these three types of hearers.

The Loudspeaker Group numbered 17. They were all students of speech therapy. One of them was Irish, one Indonesian, the rest English.

Tables

In the following tabulation of the results, the confusion matrix of G. A. Miller and P. E. Nicely (1955), also used by Householder, has been used in a simplified form.

Since the li-section has the highest percentage of correct responses, it was found convenient—for further calculations and references—to list it *before* the lɔ-section (within each group of hearers).

For easier reference the tables are numbered consecutively from I to XVI, each table giving the results of one series as interpreted by one group of hearers. Below the number of each table are indicated which group, section, and series it refers to, and also the percentage of correct responses (to the /p/-, /t/-, /k/-, and /ʔ/-signals reckoned together) of that particular series.

The symbols in the left hand column of each table indicate the signals. The letters along the top indicate the various responses of the hearers.

In the middle square of each table are couples of numbers. The first number in each couple is a percentage (rounded to a whole number). It indicates how many per cent of the /p/-, /t/-, /k/-, and /ʔ/-signals respectively were responded to as /p/, /t/, /k/, and /ʔ/ respectively. Thus Table I shows that when the hearers of the Earphone Group listened to Series I of the li-section, they interpreted 99% of the /p/-signals as /p/, 1% as /k/, and none as /t/ or as /ʔ/. Of the /t/-signals they interpreted 18% as /p/, 74% as /t/, and so forth.

The second number in each couple (printed in bold type) denotes the actual number of /p/-, /t/-, /k/-, and /ʔ/-signals respectively that were responded to as /p/, /t/, /k/, and /ʔ/ respectively.

The numbers in the right hand column of each table (also printed in bold type) denote the actual numbers of responses given (which are equivalent to the numbers of signals, since one response was required to each signal). Eight " uncertain " responses—obviously the hearer had not been able to make up his mind—are left out of account altogether.

Discussion

Table I. /p/ and /k/ are practically never misinterpreted. /ʔ/ is fairly well recognized (83%), whereas /t/ has a comparatively low percentage of correct responses (74%).

Table I
Earphone Group, li-section,
Series 1, 88·3% corr. resp.

	p		t		k		ʔ		
p	99,	**97**	0,	**0**	1,	**1**	0,	**0**	**98**
t	18,	**18**	74,	**73**	0,	**0**	7,	**7**	**98**
k	0,	**0**	0,	**0**	99,	**97**	1,	**1**	**98**
ʔ	6,	**7**	12,	**15**	0,	**0**	83,	**104**	**126**

If /ʔ/ is misinterpreted, it is taken to be /t/ (12%), or /p/ (6%), but never /k/. This *order* agrees with Householder's findings. He found that after /ɪ/ 65·5% of his hearers interpreted /ʔ/ as /t/, 22·7% interpreted it as /p/, and 13·2% as /k/ (Householder, 1956, p. 238). On the other hand, /p/ is never interpreted as /ʔ/, and /k/ very rarely so. /t/, however, is sometimes interpreted as /ʔ/ (7%). Thus, after /i/ the difference between /ʔ/ and /p/, and between /ʔ/ and /k/, seems to be fairly clear, auditively. The difference between /ʔ/ and /t/ seems to be less marked. This is to some extent borne out by Tables II, III, and IV, where the lowest percentage of correct responses is found when choice had to be made between /ʔ/ and /t/ (90·8%).

Table II
Earphone Group, li-section,
Series 2, 96·7% corr. resp.

	p		ʔ		
p	96,	162	4,	6	168
ʔ	3,	5	97,	163	168

Table III
Earphone Group, li-section,
Series 3, 90·8% corr. resp.

	t		ʔ		
t	92,	155	8,	13	168
ʔ	11,	18	89,	150	168

Table IV
Earphone Group, li-section,
Series 4, 93·2% corr. resp.

	k		ʔ		
k	94,	158	6,	10	168
ʔ	8,	13	92,	155	168

Moreover, /k/ is never confused with /t/ and very seldom with /p/. Although /p/ is never interpreted as /t/, /t/ is sometimes interpreted as /p/.

Thus Table I, supported by TABLES II, III, and IV, gives us this figure, where the unbroken lines denote " no confusion " or " rare confusion ", and the dotted lines " relatively frequent confusion ": —

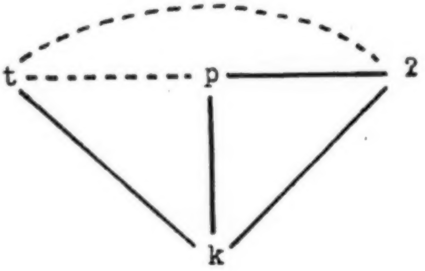


TABLE V. /p/ is well recognized (91%), and so is /t/ (90%), whereas /k/ and /ʔ/ have comparatively low percentages of correct responses (71% and 65%).

If /ʔ/ is misinterpreted, it is first of all taken to be /p/ (15%), then /k/ (10%) and /t/ (9%). Householder found that after /ɑ/, /ʔ/ was interpreted as /p/ by 50·5%, as /t/ by 28·3%, and as /k/ by 21·2% of his hearers (ibid., p. 238). So far the results in Table V are borne out by TABLES VI, VII, and VIII, where the highest

TABLE V
Earphone Group, lɔ-section,
Series 1, 78·3% corr. resp.

	p	t	k	ʔ	
p	91, 89	4, 4	2, 2	3, 3	98
t	0, 0	90, 87	0, 0	10, 10	97
k	5, 5	11, 11	71, 68	13, 12	96
ʔ	15, 19	9, 11	10, 13	65, 81	124

TABLE VI
Earphone Group, lɔ-section,
Series 2, 83·6% corr. resp.

	p	ʔ	
p	87, 146	13, 22	168
ʔ	20, 33	80, 135	168

TABLE VII
Earphone Group, lɔ-section,
Series 3, 92% corr. resp.

	t	ʔ	
t	92, 153	8, 13	166
ʔ	8, 13	92, 155	168

TABLE VIII
Earphone Group, lɔ-section,
Series 4, 91·1% corr. resp.

	k	ʔ	
k	94, 158	6, 10	168
ʔ	12, 20	88, 148	168

percentage of correct responses is found when choice had to be made between /ʔ/ and /t/ (92%), and the lowest when choice had to made between /ʔ/ and /p/ (83·6%). But although /ʔ/ is comparatively often interpreted as /p/ (15%), /p/ is not often interpreted as /ʔ/ (3%). And if /t/ is misinterpreted, it is taken to be /ʔ/.

On the whole, the responses to the lɔ-section present a more blurred picture than do the responses to the li-section. Its most prominent features are the high percentages of correct responses to the /p/-signals and the /t/-signals.

Compared with Table I, the percentage of correct responses to the /ʔ/-signals in Table V is somewhat lower (65% in Table V, 83% in Table I), and the percentages for /t/ and /k/ have become nearly inverted: 90% : 71% in Table V as opposed to 74% : 99% in Table I.

TABLE IX. Compared with Table I, Table IX has a considerably lower percentage of correct responses. The general outlines of the two are largely parallel:— /p/ and /k/ are easily recognized, and the percentages of correct responses to /ʔ/ and /t/ are lower. But the *difference* between the percentages for /p/ and /k/ on one side and for /ʔ/ and /t/ on the other is much more marked in Table IX than in Table I, and

TABLE IX
Loudspeaker Group, li-section,
Series 1, 67·1% corr. resp.

	p	t	k	ʔ	
p	97, 115	3, 3	0, 0	1, 1	119
t	42, 50	43, 51	4, 5	11, 13	119
k	0, 0	2, 2	92, 109	7, 8	119
ʔ	30, 46	23, 35	3, 5	44, 67	153

this is probably the most conspicuous feature of Table IX.

If /ʔ/ is misinterpreted, it is often taken to be /p/ (30%) and /t/ (23%), but very seldom /k/ (3%). /p/ is practically never interpreted as /ʔ/ (1%), and /k/ comparatively seldom so (7%). /t/ is sometimes taken to be /ʔ/ (11%). Except for the many /ʔ/-signals that are interpreted as /p/, these results are in conformity with Tables X, XI, and XII, where the lowest percentage of correct responses is found when choice had to be made between /t/ and /ʔ/ (62·5%).

/k/ is never confused with /p/, and seldom with /t/. Although /p/ is seldom interpreted as /t/ (3%), /t/ is very often interpreted as /p/ (42%).

Table X
Loudspeaker Group, li-section,
Series 2, 88% corr. resp.

	p		ʔ		
p	95,	194	5,	10	204
ʔ	19,	39	81,	165	204

Table XI
Loudspeaker Group, li-section,
Series 3, 62·5% corr. resp.

	t		ʔ		
t	64,	131	36,	73	204
ʔ	39,	80	61,	124	204

Table XII
Loudspeaker Group, li-section,
Series 4, 84·3% corr. resp.

	k		ʔ		
k	83,	169	17,	35	204
ʔ	14,	29	86,	175	204

Table XIII. The general outline corresponds to that of Table V, only the percentages of correct responses are much lower. The highest percentages are found for /t/ (49%) and /p/ (47%), whereas /k/ and /ʔ/ have percentages that are lower than pure chance would be expected to give: 22% and 16%.

Table XIII
Loudspeaker Group, lɔ-section,
Series 1, 32·2% corr. resp.

	p	t	k	ʔ	
p	47, 56	15, 18	19, 23	18, 22	119
t	18, 21	49, 58	10, 12	24, 28	119
k	21, 25	17, 20	22, 26	40, 47	118
ʔ	48, 73	14, 22	22, 34	16, 24	153

If /ʔ/ is misinterpreted, it is first of all taken to be /p/ (48%) and then /k/ (22%) and /t/ (14%). This is in conformity with Tables XIV, XV, and XVI, where the highest percentage of correct responses is found when choice had to be made between /t/ and /ʔ/ (58·6%), and the lowest when choice had to be made between /p/ and /ʔ/ (43·4%). But, as in Table V, we find that although /ʔ/ is often taken to be /p/ (48%), /p/ is not so often taken to be /ʔ/ (18%).

Table XIV
Loudspeaker Group, lɔ-section,
Series 2, 43·4% corr. resp.

	p	ʔ	
p	49, 99	51, 105	204
ʔ	62, 126	38, 78	204

Table XV
Loudspeaker Group, lɔ-section,
Series 3, 58·6% corr. resp.

	t	ʔ	
t	56, 115	44, 89	204
ʔ	39, 80	61, 124	204

Table XVI
Loudspeaker Group, lɔ-section,
Series 4, 54·4% corr. resp.

	k	ʔ	
k	55, 112	45, 92	204
ʔ	46, 94	54, 110	204

Conclusions

The total number of responses given by the Earphone Group is 2,849 (Tables I - VIII), of which 2,534 = 88·9% are correct. The total number of responses given by the Loudspeaker Group is 3,467 (Tables IX - XVI), of which 2,102 = 60·6% are correct.

Although the two groups were two sets of persons, this result strongly indicates that correct interpretation largely depends on the way in which the signals are (re)produced and conveyed to the hearers. The hearers of the Earphone Group listened one at a time. The earphones were tight-fitting, and nearly all noise from outside was eliminated. The hearers of the Loudspeaker Group were sitting in an ordinary room, at different distances from, and angles to, the apparatus, with no special elimination of noise.

A comparison between Table I and Table IX, Table V and Table XIII shows that less favourable listening conditions as a rule have the most damaging effect on sounds that have the lowest percentage of correct responses under favourable conditions. Thus, when listened to by earphones 99% of the /lip/-signals of Series 1 (Table I) were correctly responded to. When listened to by loudspeaker, the percentage falls only very little, to 97% (Table IX). The corresponding percentages for the /liʔ/-signals present a considerable fall: from 83% to 44%. In the same way the percentages for the /lɔp/-signals drop from 91% to 47%, and the percentages for the /lɔʔ/-signals from 65% to 16% (Tables V and XIII).

The total number of responses in the li-section (Tables I - IV, IX - XII) is 3,162, of which 2,614 = 82·7% are correct. The total number of responses in the lɔ-section (Tables V - VIII, XIII - XVI) is 3,154, of which 2,022 = 64·1% are correct.

This indicates that correct interpretation of unreleased voiceless plosives largely depends on the quality of the preceding vowel, as is also shown by Householder, who obtained 72% correct responses after /ɪ/ and 46% after /ɑ/ (ibid., p. 240).

Especially to the interpretation of the /p/- and the /k/-signals the preceding vowel is of importance. Of the 589 /lip/-signals, 568 = 96·4% had correct responses. Of the 589 /lɔp/-signals, only 390 = 66·2% had correct responses. Of the 589 /lik/-signals, 533 = 90·5% had correct responses. Of the 586 /lɔk/-signals, only 364 62·1% had correct responses.

The vowel is of less, though considerable, importance to the correct interpretation of the /ʔ/-signals. Of the 1,395 /liʔ/-signals, 1,103 = 79·1% had correct responses. Of the 1,393 /lɔʔ/-signals, 855 = 61·4% had correct responses.

On the average the vowel is of little importance to the correct interpretation of the /t/-signals. Of the 589 /lit/-signals, 410 = 69·6% had correct responses. Of the 586 /lɔt/-signals, 413 = 70·5% had correct responses. The percentage of correct responses is actually higher after /ɔ/ than after /i/, a feature that is particularly prominent when choice had to be made among 4 alternatives (Tables I and IX).

As to the possibility of identifying an unreleased glottal stop auditively, the experiment gives the following results:—

1. *The Earphone Group*

In the cases where the hearers had to choose between two possibilities (Tables II - IV, VI - VIII), the 1,008 /ʔ/-signals had 906 = 89·9% correct responses, varying from 97% (/liʔ/ Table II) to 80% (/lɔʔ/ Table VI). In the same cases 74 = 7·4% out of the 1,006 /p/-, /t/-, and /k/-signals were interpreted as /ʔ/, varying from 13% (/lɔp/ Table VI) to 4% (/lip/ Table II).

In the cases where the hearers had to choose out of four possibilities (Tables I and V), the 250 /ʔ/-signals had 185 = 74% correct responses, varying from 83% (/liʔ/ Table I) to 65% (/lɔʔ/ Table V). In the same cases 33 = 5·6% out of the 585 /p/-, /t/-, and /k/-signals were interpreted as /ʔ/, varying from 13% (/lɔk/ Table V) to 0% (/lip/ Table I).

These figures tend to show that an unreleased glottal stop is recognized auditively, after one hearing, with considerable accuracy (89·9% if there are 2 alternatives ; 74% if there are 4 alternatives), provided that the listening conditions are of the standard offered by the two types of earphones mentioned. Under the same conditions unreleased /p/, /t/, and /k/ are not often mistaken for /ʔ/ (7·4% if there are 2 alternatives ; 5·6% if there are 4 alternatives).

As has already been pointed out, the percentage of correct responses to the /ʔ/-signals is higher after /i/ than after /ɔ/. If choice had to be made between 2 possibilities, 468 = 92·9% of the 504 /ʔ/-signals were interpreted correctly after /i/, 438 = 86·9% out of 504 after /ɔ/. If choice had to be made among 4 possibilities, the corresponding figures are 104 = 82·5% out of 126 after /i/, and 81 = 65·3% out of 124 after /ɔ/.

2. *The Loudspeaker Group*

In the cases where the hearers had to choose between 2 possibilities (Tables X - XII, XIV - XVI) the 1,224 /ʔ/-signals obtained 776 = 63·4% correct responses, varying from 86% (/liʔ/ Table XII) to 38% (/lɔʔ/ Table XIV). In the same cases 404 = 33% out of the 1,224 /p/-, /t/-, and /k/-signals were interpreted as /ʔ/, varying from 51% (/lɔp/ Table XIV) to 5% (/lip/ Table X).

In the cases where the hearers had to choose out of 4 possibilities (Tables IX and XIII), the 306 /ʔ/-signals had 91 = 29·7% correct responses, varying from 44% (/liʔ/ Table IX) to 16% (/lɔʔ/ Table XIII). In the same cases 119 = 16·7% of the 713 /p/-, /t/-, and /k/-signals were interpreted as /ʔ/, varying from 40% (/lɔk/ Table XIII) to 1% (/lip/ Table IX).

These figures suggest that under the listening conditions of the Loudspeaker Group the possibility of identifying an unreleased glottal stop auditively after one hearing seems—on the whole—to be only slightly better than pure chance.

The percentages of correct responses to the /ʔ/-signals are markedly higher, however, after /i/ than after /ɔ/. If choice had to be made between 2 possibilities, 464 = 75·8% out of the 612 /ʔ/-signals were interpreted correctly after /i/, 312 = 51% out of 612 after /ɔ/. If choice had to be made among 4 possibilities, the corresponding figures are: 67 = 43·8% out of 153 after /i/, and 24 = 15·7% out of 153 after /ɔ/.

References

Householder, F. W., Jr. (1956). Unreleased PTK in American English. In *For Roman Jakobson, Essays on the occasion of his sixtieth birthday*, ed. Morris Halle (The Hague), 235.

Miller, G. A. and Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *J. acoust. Soc. Amer.*, 27, 338.

NEW JOURNAL OF AUDITORY RESEARCH

The board of editors announces the founding of The Journal of Auditory Research, an interdisciplinary non-profit quarterly devoted to the scientific study of hearing. Publication will cover the fields of psycho-acoustics, otology, audiology, neurophysiology of audition, speech and communications, auditory aspects of human engineering, musicology, instrumentation for hearing research, and all other aspects of audition. Special policies will include quick but thorough editing, rapid publication and lowest possible cost to subscribers.

Dr. J. Donald Harris, of the Medical Research Laboratory, New London, Connecticut, will be Editor of the new journal, and Dr. James Jerger, of Northwestern University, Evanston, Illinois, will be Associate Editor. The Editorial Policy Board consists of Drs. Norton Canfield, Raymond Carhart, Stacy Guild, Henry L. Haines, Fred Kranz, Alvin M. Liberman and E. Glen Wever. The journal will be published by the C. W. Shilling Auditory Research Center, Inc., of Groton, Connecticut. Initial publishing costs have been defrayed by a grant from the Beltone Institute for Hearing Research.

Manuscripts and preliminary reports on all aspects of hearing are invited. They should be addressed to the Editor, 348 Long Hill Road, Groton, Connecticut, U.S.A. The advance subscription will be $3.00 and the regular subscription $5.00.

PUBLICATIONS RECEIVED

Abstracts of English Studies, 2 (1959), 9-12 ; 3 (1960), 1-3.

Acta Linguistica Academiae Scientiarum Hungaricae, 9 (1959), 1/2, 3/4.

American Annals of the Deaf, 104 (1959), 5.

Behavioral Science, 4 (1959), 4 ; 5 (1960), 1.

ETC., 16 (1958), 1 ; 16 (1959), 3-4.

Journal of Speech and Hearing Research, 2 (1959), 3-4 ; Monograph Supplement 5, Sept. 1959 ; 3 (1960), 1.

Leuvense Bijdragen, 49 (1960), 1/2.

Mechanical Translation, 5 (1958), 3.

Methodos, 11 (1959), 41.

Problems of Linguistics (Moscow), 8 (1959), 5-6 ; 9 (1960), 1.

Revue de Linguistique (Bucarest), 4 (1959), 1.

Slovo a Slovesnost, 20 (1959), 4.

Studia Romanica et Anglica Zagrabiensia, 7 and 8 (1959).

Volta Review, 61 (1959), 10 ; 61 (1960), 1, 3.

Coteanu, I., Iordan, I., Rosetti, A. (eds.) (1959). Recueil d'Etudes Romanes (Academia Republicii Populare Romine, Bucarest).

Iordan, I., Petrovici, E., Rosetti, A. (eds.) (1957). Mélanges Linguistiques (Academia Republicii Populare Romine, Bucarest).

Iordan, I., Petrovici, E., Sala, M. (eds.) (1958). Contributions Onomastiques (Academia Republicii Populare Romine, Bucarest).

Kaplan, Harold M. (1960). Anatomy and Physiology of Speech (McGraw-Hill, New York).

Rosetti, A. (ed.) (1958). Omagiu lui Iorgu Iordan (Academia Republicii Populare Romine, Bucarest).

Rosetti, A. (ed.) (1958). Fonetica si Dialectologie, Vols. I and II (Academia Republicii Populare Romine, Bucarest).

Rosetti, A. (ed.) (1959). Recherches sur les Diphtongues Roumaines (Academia Republicii Populare Romine, Bucarest and Einar Munksgaard, Copenhagen).

Vachek, J. and Firbas, J. (1959). Phonic Analysis of Present-day English (Statni Pedagogicke Nakladatelstvi, Prague).

Research Needs in Speech Pathology and Audiology (American Speech and Hearing Association, 1959).

COMMUNICATION OF VERBAL MODES OF EXPRESSION*

IRWIN POLLACK, HERBERT RUBENSTEIN AND ARNOLD HOROWITZ**
Operational Applications Office, Air Force Command and Control Development Division, Bedford, Massachusetts

Talkers were instructed to read neutral sentences and ' sound happy ', or ' sound bored ', etc. Listeners attempted to identify the intended mode of expression drawing their responses from a limited number of alternatives. Results are presented showing how the identification of modes of expression is affected by: (1) number of response alternatives, (2) noise, (3) whispering, and (4) temporal sampling. Reasonably high levels of performance may be achieved under conditions of reduced acoustic information.

Because of severe security and/or bandwidth requirements, future military voice communication systems will require extensive speech processing. In addition to the criterion of speech intelligibility, other criteria of performance will be required for these systems. Factors taken for granted with unprocessed speech communication systems, e.g., talker recognition (Ochiai, 1959 ; Pollack, Pickett, and Sumby, 1954), naturalness of speech, recognition of nuances of pitch, duration, and stress (Denes, 1959 ; Fry, 1958 ; Lieberman, 1960) may be seriously impaired in highly processed systems (Shearme and Holmes, 1959).

The present study explores one facet of performance beyond intelligibility—recognition of the talker's verbal expression. We wished to determine whether listeners can recognize whether a talker has been instructed to ' sound happy ', or ' sound bored ', etc., in the reading of a neutral sentence. In this exploration, we have restricted our examination to conventional speech channels without special processing. In co-operation with other groups, we hope to be able to extend this work to advanced processing systems.

* *This is Technical Note 60-24 of the Air Force Command and Control Development Division. This research is in support of Project 7684, Communications in Noise, of the Air Research and Development Command Programme in Human Performance. The authors wish to thank Mr. William H. Sumby who served as the fourth talker.*

** *Now at The Franklin Institute, Philadelphia, Pennsylvania.*

TABLE 1

INITIAL RATING	FINAL RATING	MODE OF EXPRESSION
1	2	pedantic
2	1	boredom
3	4	objective question
4	5	confidential communication
5	8	objective statement
6	3	fear
7	10*	uncertainty
8	6	disbelief
9	7	happiness
10	9*	impatient repetition
11	12*	sarcasm
12	11*	threat
13	13*	approval
14	14*	disgust
15	15*	surprise
16	16*	anger

Modes of expression employed in this study.
The constant-ratio rule was successively applied upon the initial 16×16 confusion matrix to achieve the final ratings.

PROCEDURE

The literature on traits and character recognition was first searched for lists of expressive modes. Each of the four talkers independently recorded several neutral sentences in as many different modes of expression as he could interpret.

The 16 modes which were best recognized in preliminary tests with laboratory personnel were selected for future experimentation. These modes are listed in Table 1.

Each of four talkers recorded the two sentences, " The lamp stood on the table " and " His friend is coming by train " in each of the 16 modes. The modes were presented in randomized order. The order of presentation of the 16 modes was different for each talker. Before the initiation of the study, both the talkers and listeners were naive as to the experimental recognition or production of modes of verbal expression. The talkers were not trained actors. Only broad restrictions were placed upon the talker's mode of expression. Extra sounds (coughs, grunts, long pauses, etc.) were prohibited. The talkers monitored their voices with a VU meter attempting to achieve either a fixed VU setting upon the stressed words of the sentence (3 talkers) or to achieve a constant VU setting over the sentence (1 talker).

The recordings were played back to each of three crews of six listeners each. Each listener was equipped with the same list of modes employed by the talkers. The

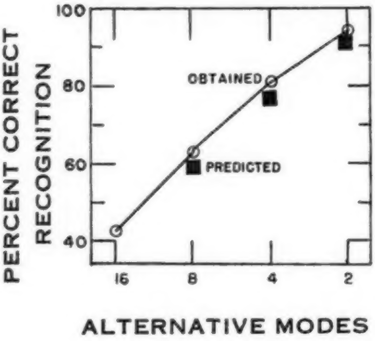


Fig. 1. Test of the constant-ratio rule. The ordinate represents the average per cent correct recognitions. The abscissa is the number of response alternatives available. The circles represent obtained results; the squares represent predictions by the constant-ratio rule of Clarke. Each point represents the average of 2300 observations: 32 observations by each of 18 listeners with each of 4 talkers.

listener was not informed whether his responses were correct or incorrect. After the listeners were exposed to the entire set of materials, the same materials were played back, but with only eight response alternatives available per item; then with four response alternatives per item; and, finally, with two response alternatives per item. All tests were carried out over a loudspeaker system in a quiet room. No additional noise, other than residual system noise, was employed.

Test of constant-ratio rule

The results of the initial listening tests are presented in Table I and Fig. 1. The left column of Table 1 presents the rank-ordering of the 16 modes of expression in the initial tests with 16 alternatives. Fig. 1 presents the average correct recognition score as a function of the number of response alternatives. The circles represent the observed scores. The squares represent the corresponding scores predicted by the constant-ratio rule of Clarke (1957). This rule, in essence, states that the ratio of any two conditional response confusions is independent of the other members of the confusion set. Predictions for *n* response alternatives were based on the obtained result for 2*n* response alternatives. Specifically, predictions for 2, 4, and 8 response alternatives were based upon the obtained results with 4, 8, and 16 response alternatives respectively and Table 2 presents the results of application of the constant-ratio rule. The average discrepancy between the average obtained and predicted correct percentage (line 3) ranges from 1 to 4%. The direction of the discrepancy is consistent with the assumption that the listening crew became more experienced in proceeding from the 2*n*- to the *n*-alternative tests. The mean difference, sign ignored, between all of the predicted and obtained entries (line 1) ranged from 3 to 5%. The mean difference, sign ignored, for the correct entries only (line 2) was about 6·5%. The accuracy of prediction is

TABLE 2

	16	8	4
1. Mean absolute difference, predicted vs. obtained, all entries	2·9%	4·0%	4·6%
2. Mean absolute difference, diagonal entries only	6·5%	6·4%	6·5%
3. Obtained correct minus predicted correct	3·3%	3·9%	0·8%
4. % of differences greater than 0·10	6·2%	10·9%	0·0%

Prediction by the constant-ratio rule from $2n$- to n-alternative tests.

considered to be good. Analyses of variance performed on the results of 16 modes showed that: talker, modes, groups, and sentences were statistically significant as were various higher order interactions thereof.

Selection of eight ' best ' modes.

With some assurance that the constant-ratio rule was applicable to modes of expression, we attempted to select the most discriminable set of eight modes from the initial sixteen. The following steps were performed: From the 16 × 16 confusion matrix, two scores were obtained for each mode of expression: the percentage of trials in which each stimulus mode was correctly identified ; and the percentage of trials in which each of the alternative 16 responses was correctly assigned. The two scores were averaged. It was necessary to obtain both scores because some of the response alternatives were employed more often than others. (The 16 modes in Table 1 are first ordered in terms of the average scores obtained in the 16 × 16 tests.) The mode with the lowest average correct score was eliminated, as indicated by 16* in the second column. The initial 16 × 16 matrix was reduced to a 15 × 15 matrix with the eliminated entries redistributed back into the new matrix according to the constant-ratio rule. For example, if modes *A* and *B* were highly confused with respect to each other, but not with respect to other modes, elimination of mode *B* might bring about a striking improvement in mode *A*. The procedure was repeated until only eight modes of expression remained. The final rank-order of the eight modes is indicated in the second column of Table 1. Entries with asterisks were eliminated before the final selection.

Fig. 2 presents the change in performance as a result of successive reductions in the size of the number of response alternatives. The eight modes finally selected are indicated by dots. The crosses are associated with modes which failed to remain after the final selections, i.e., with asterisks in Table 1. The bars are drawn through the means of the eight dots. It is important to note that only the points above the abscissa value 0 in Fig. 2 represent experimental observations. Thereafter, successive changes are introduced by application of the constant-ratio rule.

The successive reductions in response alternatives serve to increase the average correct reception score from about 55% to 75% correct ; and also to reduce the spread among the separate modes.

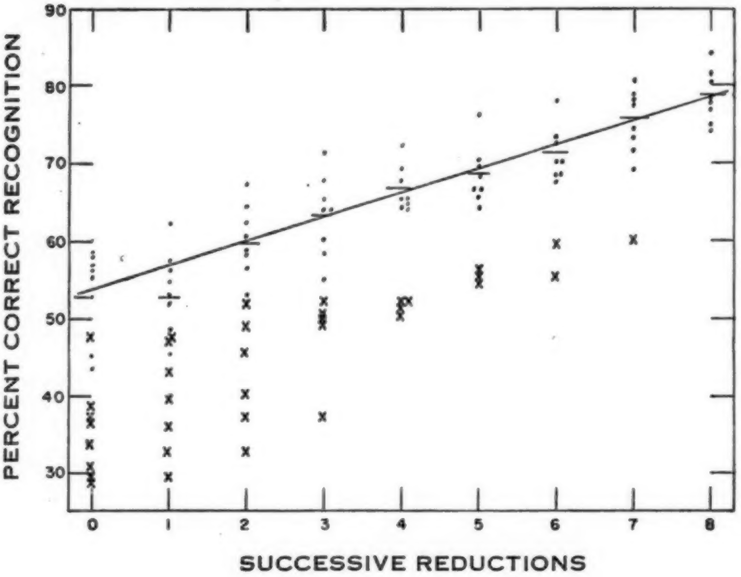


Fig. 2. Improvement in average recognition score with successive reductions of the size of the ensemble. The scores associated with the eight modes that were finally selected are plotted as dots; the scores associated with the modes that were eventually rejected are presented as crosses. The bar represents the mean score for the selected eight modes only.

Mode vs sentence recognition

The following tests attempt to compare the recognition of expressive modes with the recognition of short sentences. The sentences are presented in Table 3. Each sentence is a declarative sentence consisting of three monosyllabic content words and three monosyllabic function words.

Experimental Procedure.

The procedure differed from the preliminary experiment in the following details: Only eight modes of expression were employed in all tests. Live voice was employed. The talkers, however, were not visible to the listeners. A trial-block was defined as 24 presentations. In the identification of modes of expression, each of the eight modes was read on three occasions in a randomized order; and only a single sentence was employed for each trial-block. In the identification of sentences, each of the eight sentences was read on three occasions in a randomized order; and only a single mode

TABLE 3

The lamp stood on the desk
They have bought an old car
He will work hard next term
His friend came home by train
She parked near the street light
We talked for a long time
John found him at the phone
You have seen my new house

Sentences employed in the tests.

was employed for each trial-block. The listeners employed a set of eight push-buttons. The correct answer was available after each presentation. The equipment permitted the determination of the complete input-output matrix in real time. The materials were read against a white noise (100 - 6800 c.p.s.) background over earphones. The speech to noise (S/N) ratios represent VU readings upon the speech and true r.m.s. readings upon the noise.

Results.

The results are presented in Fig. 3 in terms of the average percentage correct recognition of the eight modes of expression (circles) and of the eight sentences (squares), as a function of S/N ratio. The small graphs present the results for the individual talkers. The first of these graphs represents the performance of the talker who attempted to achieve a constant VU setting over the entire sentence.

The difference between the recognition of modes and sentences is small. For each of the graphs, mode recognition is somewhat superior at extremely poor S/N ratios; whereas, sentence recognition is superior at more favourable S/N ratios. In this connection it should be noted that the sentences, unlike the modes, were not selected for high discriminability. Nevertheless the comparison suggests that mode identification is possible under conditions where only a small segment of the sentence is recognized.

WHISPERING TESTS

We next turned to some factors which might be responsible for recognition of modes of verbal expression. We were first concerned with the role of voicing pitch. In order to examine its role, we employed whispered speech which reduces, although it does not entirely eliminate, speech pitch characteristic (Meyer-Eppler, 1957). Two of the original four talkers were employed; the results in the whispered tests for each talkei are presented in Fig. 4 along with his results for the normally-voiced tests.

The equivalence of results at low S/N ratios disappears in favour of the voiced tests at higher S/N ratios. Nevertheless, reasonably high scores may be obtained with

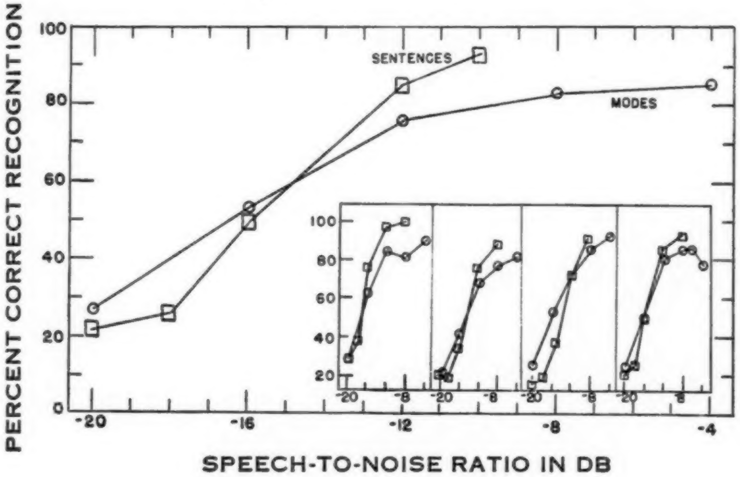


Fig. 3. Mode and sentence identification in noise. The insert graphs present the results of the individual talkers. Each point is the main graph represents 4600 observations: 96 observations by each of 12 listeners with each of 4 talkers.

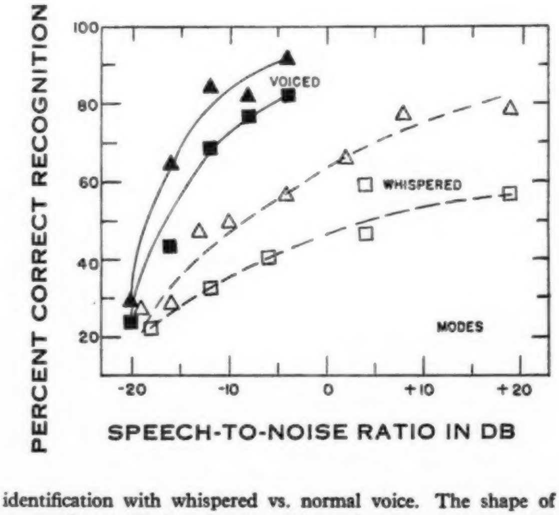


Fig. 4. Mode identification with whispered vs. normal voice. The shape of the points codes the two talkers employed. Each point for whispered voice represents 576 observations: 96 observations by each of 6 listeners.

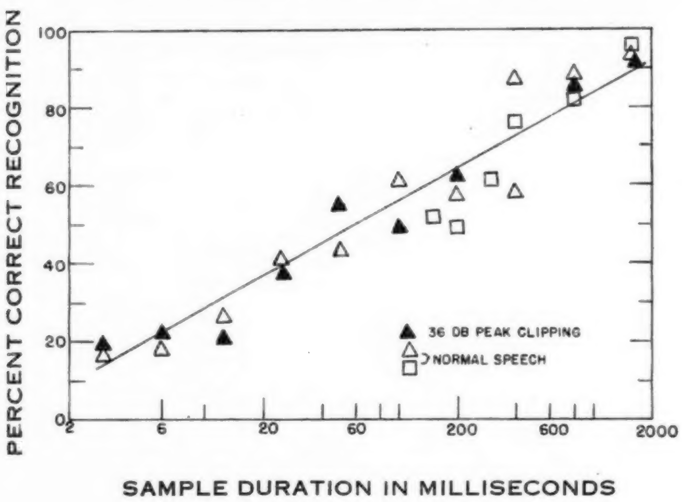


Fig. 5. The effect of duration of utterance upon the identification of modes of expression. The shape of the points codes the two talkers employed. Each point is based upon 360 observations: 72 observations by each of 5 listeners.

whispered speech. Thus, we conclude that voice pitch alone is not critical for the identification of modes of expression.

Temporal factors: single samples

We next attempted to examine temporal factors in the recognition of modes of verbal expression. Ideally, we would have liked to have controlled the temporal stress pattern as might be accomplished upon speech synthesizers such as the Pattern Playback of the Haskins Laboratory. Since we did not have this capability, we decided to examine the effect of the duration of utterance. A voice-operated relay was tripped by the talker's voice. In turn, a gate was opened for a pre-determined interval. The effect of the gated interval is presented in Fig. 5 for two of the initial four talkers (open figures). In addition, results obtained by one talker at 36 db peak-clipping are presented.

The per cent correct mode recognition is approximately linear with the log of the sample duration. Mode recognition is possible with extremely short samples. About 50% correct recognition is obtained with a 60 msec. sample. Severe peak-clipping does not destroy this recognition. It would again appear that mode recognition is possible under minimal acoustic information.

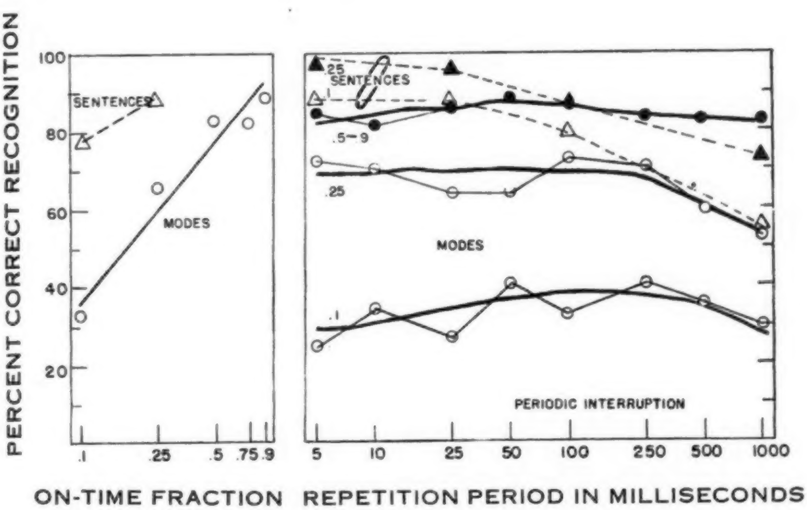


Fig. 6. The effect of periodic sampling upon mode identification. The graph on the left presents the results of the graph on the right averaged over repetition rate. Each point in the right graph represents 240 observations: 24 observations by each of 5 listeners for each of 2 talkers.

Temporal factors: periodic sampling

We next attempted to examine the effect of periodic sampling upon sentence and mode recognition. The talker's utterances were led through an electronic switch (Grason-Stadler 829) in which the on-fraction—the percentage of time that speech was available—and repetition period were independently varied.

The average results obtained with two talkers are presented in Fig. 6. On the left, the results, averaged over a range of repetition periods (5-1000 msec.), are presented for both sentence and mode recognition as a function of on-time fraction. It is clear that periodic sampling is considerably more deleterious to mode recognition than to sentence recognition. Again, per cent correct mode recognition is approximately linear with the log of the sample duration. Assuming a total duration of the sentence at 1·6 sec., the results of periodic sampling and a single sample are quite comparable.

On the right, the results are presented as a function of the repetition period for on-time fractions of 0·1 and 0·25 for both mode and sentences. In addition, results are presented for mode recognition averaged over on-time fractions of 0·5, 0·75 and 0·9. Repetition period operates somewhat differently upon modes and sentence recognition. Sentence recognition falls off with longer repetition periods. Mode recognition appears to be less sensitive to the repetition period, although a repetition period of 100 msec. appears to be an optimal sampling period.

Conclusion

High levels of correct recognition may be achieved for a defined small number of modes of verbal expression for a defined small number of talkers by a listening crew familiar with the talkers' modes of expression. Reasonably high levels of recognition may be achieved with sharply reduced acoustical information (low S/N ratios, short samples, or whispered speech).

References

Allport, G. W. and Vernon, P. E. (1933). Studies in Expressive Movement (New York).

Clarke, F. R. (1957). Constant-ratio rule of confusion matrices in speech communications. *J. acoust. Soc. Amer.*, **29**, 715.

Denes, P. (1959). A preliminary investigation of certain aspects of intonation. *Language and Speech*, **2**, 106.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, **1**, 126.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *J. acoust. Soc. Amer.*, **32**, 451.

Meyer-Eppler, W. (1957). Realization of prosodic features in whispered speech. *J. acoust. Soc. Amer.*, **29**, 104.

Ochiai, Y. (1959). Phoneme and voice identification studies using Japanese vowels. *Language and Speech*, **2**, 132.

Pollack, I., Pickett, J. M. and Sumby, W. (1954). On the identification of speakers by voice. *J. acoust. Soc. Amer.*, **26**, 403.

Shearme, J. N. and Holmes, J. N. (1959). An experiment concerning the recognition of voices. *Language and Speech*, **2**, 123.

FREQUENCY STUDIES OF ENGLISH CONSONANTS*

WILLIAM S-Y. WANG AND JOHN CRAWFORD
University of Michigan

In this study an explanation was sought for the disagreement among the various frequency counts which have been made of English consonants. The data for a set of ten different frequency counts were converted to IPA symbols and compared by means of the coefficient of linear correlation. It was found that the relative frequency of consonants in English is not seriously affected by the style of literary content or by the dialect of the sample and that a relatively small sample yields typical values. Differences in the general type of corpus (dictionary or running texts) and in transcription, however, cause significant discrepancies among the various studies. It is concluded that higher order frequency data are probably more relevant to mechanical speech recognition than the first order data considered in this paper.

MECHANICAL SPEECH RECOGNITION AND LANGUAGE STATISTICS

In a series of articles dealing with the problems of mechanical speech recognition, Fry and Denes (1957, 1958) discussed the results of an experiment which involved the use of language statistics. The experiment showed that with an identical input, their mechanical recognizer achieved a considerably higher articulation score when a limited amount of language statistics was programmed in the recognition operation.

The input to the recognizer was a series of approximately 500 phone tokens consisting of 12 English phone types. The series was composed of 140 words, each of which was bounded by spaces (periods of silence) which could be recognized with perfect accuracy. The statistical information they used was the 2nd order probability[1], i.e., the probability that the phone is *j* given that the immediately preceding phone is *i*.

* *This study was supported by the Office of Scientific Research of the U.S. Air Force under Contract AF-49-(638)-492. The authors wish to thank Professor G. E. Peterson of the University of Michigan and Miss E. Sivertsen of the University of Michigan and the Norwegian Advanced Teachers' College, Trondheim, Norway, for their help in preparing this paper.*

[1] $P_i j$ *is the conditional probability that* j *occurs immediately after a given* i; *in general,* $P_i j$ *is called the* (n + 1)*th order probability when* i *is a sequence of* n *phone tokens.*

It is obvious that if the input is an idiolect[2], rather than a closed set of isolated words, the problem which the recognizer encounters becomes immensely more difficult. In this case, both the number of phone types and the number of possible sequences are considerably larger. The sequences between spaces are usually much longer, thus increasing the likelihood of sequential errors.[3] Furthermore, with the tempo fluctuations which occur in normal speech, it is not certain that rules can be devised to distinguish consistently the spaces from the voiceless gaps of plosive consonants.[4]

The experiment by Fry and Denes demonstrates in a striking manner the possible usefulness of statistical information in the design of speech recognizers. This experiment indicates that the collection of such information about natural speech may be an important step toward the realization of mechanical speech recognition.

The first order probability of English phones

An elementary form of statistical information about a language is the relative frequencies of the phones, i.e., the first order probabilities. Clearly, the more complex forms of statistical information are dependent on the results of the first order probabilities, e.g.,

$$P_{ab} = \frac{n(ab)}{n(a)}$$

where:

P_{ab} = the conditional probability that *b* occurs next, given that *a* has just occurred.

$n(a)$ = the total number of *a* tokens.

$n(ab)$ = the total numbers of *ab* tokens.

Many studies of the first order probabilities of English sounds are available in the literature, some of which are listed in the appendix of this paper. However, a cursory examination of these studies shows that there are discrepancies among their results which cannot be immediately explained. It is well known that any natural language is a sample that is essentially open-ended and highly heterogeneous. Therefore, it is not surprising to find that various statistical studies of the same language should yield different results. However, it seems desirable to be able to specify the amount of

[2] *An idiolect is here defined as the set of all possible sentences of a language which may be pronounced by one speaker.*

[3] *This problem is inherent in any Markov chain approach to speech recognition. However, if the recognizer is not required to operate in real time, the problem can be obviated. Instead of considering the sequencies of probabilities unidirectionally, those portions of the sentence which are recognized acoustically with a higher degree of reliability (e.g., space, stressed vowels) may be used as "reference points". The recognition of the adjacent phones can then be based on these reference points and the probabilities used bidirectionally. Prosodic information and a stored dictionary may also be used in achieving an optimal recognition of the sentence (Peterson, 1959).*

[4] *We have examined the effectiveness of the rule "space $\geq$ 200 msec. $>$ gap" on a 30 minute sample of recorded speech. Only 5 plosives had longer gaps, but a large number of spaces did not exceed 200msec.*

agreement among the results and, if possible, to account for the differences. The comparison and evaluation of the previous results seem essential to the effective accumulation of further statistical information about consonant frequencies.

It is the purpose of this paper to examine the results of the above indicated studies in terms of the degree of correlation among them. The method adopted here is standard, viz., a computation of the coefficients of linear correlation. It was found that several of the studies have results which are very similar. Furthermore, many of the differences among the studies are explainable in terms of relatively simple factors.

The correlation coefficient

The first order probabilities of the phones of English may be presented as a list of phone types, with a probability associated with each phone type. The correlation between any pair of lists represents the total agreement between the pairs of probabilities associated with the various phone types. There are, of course, many methods available for the measurement of the relationship between each pair of lists. For the present study, we have selected the coefficient of linear correlation *r*, which is defined by the following relation (Hoel, 1954).

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n\, s_x\, s_y}$$

where:

$\bar{x}$ = the mean of list X.
$\bar{y}$ = the mean of list Y.
s_x = the standard deviation of list X.
s_y = the standard deviation of list Y.
n = the number of phone types.

Consider list X which gives the probabilities of the phone types as $x_1, x_2, \ldots, x_i, \ldots, x_n$, and list Y with the probabilities $y_1, y_2, \ldots, y_i, \ldots, y_n$. Each list contains n probabilities, and x_i and y_i are the probabilities of the ith phone type. If the values (x_i, y_i) are plotted on the Cartesian plane, the relationship between the two lists can be exhibited in the form of a scatter diagram. The measure of relationship between the two lists is independent of the choice of origin for the variables because it is a function of $(x_i - \bar{x})$ and $(y_i - \bar{y})$, so that the origin is at the point $(\bar{x}, \bar{y})$. Also, r is independent of the length of the lists and of the scale of measurement used for x and y, because of the factors n and $s_x\ s_y$ respectively in the denominator. It may be noted that r satisfies the inequality $-1 \leq r \leq 1$, and if for every pair (x_i, y_i), we have $x_i = y_i$, then $r = 1$. The points will lie on a straight line if there is some constant k such that for every pair (x_i, y_i) we have $x_i = ky_i$, in particular, if for every pair (x_i, y_i) we have $x_i = -y_i$, then the points lie on the line $x = -y$ and $r = -1$.

Table 1

AUTHOR	DIALECT	SAMPLE SOURCE	NUMBER OF CONSONANTS	PER CENT OF CONSONANTS
TRN	Southern British (D. Jones)	Dictionary	15,459	73.74
FOW	Wisconsin	Modern prose	10,194	67.96
CAR	Rhode Island	Modern plays	10,784	66.07
HAY	General American	Lectures	41,412	63.95
WHI	New England	Modern prose	6,371	63.71
DEW	Standard Dictionary (Funk and Wagnalls)	Newspapers	235,025	63.06
VOE	General American	Radio announcements	409,506	61.99
FRE	Unspecified	Telephone conversations	135,548	61.03
FRY	Southern British (D. Jones)	Phonetic reader	10,305	60.61
TOB	General American (Kenyon and Knott)	Telephone conversations	133,460	60.09

Frequency studies compared in this paper.

Interpretation of previous studies

In Table 1, a brief description is given of the ten studies which were examined for this paper. A list of these ten studies is given in the Appendix. In the tables the authors' names are abbreviated to the first three letters. Two major modifications were made of the 10 studies listed in Table 1. One is that the present comparison is restricted only to consonants. The other is that certain normalizations of the frequency lists were undertaken to make the lists more comparable. Thus both the number and the percentage of consonants in the total sample, as presented in Table 1, differ from the figures given in the original studies.

The comparison is restricted to the consonant system for the following reasons: (1) As will be seen later, it is relatively simple to establish a comparison among the consonant systems of the various studies. There is a great deal more discrepancy among the various studies in the vowel data than in the consonant data. (2) In general, it is difficult to distinguish the discrepancies among the vowel systems which are due to transcriptional differences from those which reflect differences in phonemic distinctions. (3) Dialectal differences are considerably more complex among vowels than among consonants. Of course, a complete comparison of data on different dialects should not only include the vowels, but also the prosodic structures.

Table 2 presents the rank ordering of the consonants as given in the original studies. The symbols have been normalized to those of the IPA. It can be seen that there are three consonants which are not regularly represented in the various lists. These are /tʃ/, /dʒ/, and /ʍ/. In the present analysis these have all been treated as sequences of consonants, i.e., /t/ and /ʃ/, /d/ and /ʒ/, /h/ and /w/ respectively. This operation

Table 2

	TRN	FOW	CAR	HAY	WHI	DEW	VOE	FRE	FRY	TOB
1.	t	t	j	n	r	n	n	t	n	t
2.	l	n	t	t	n	t	t	n	t	n
3.	s	r	n	r	t	r	r	r	d	d
4.	k	j	w	s	d	s	d	l	s	s
5.	n	w	r	l	s	d	s	d	l	l
6.	r	s	h	ð	l	l	l	ð	ð	ð
7.	p	d	s	d	ð	ð	ð	s	m	r
8.	d	l	l	k	m	z	m	m	k	m
9.	m	k	m	m	z	m	k	k	r	k
10.	b	ð	d	z	v	k	f	w	w	w
11.	g	z	k	v	h	v	z	z	z	j
12.	f	m	ð	p	w	w	b	j	v	z
13.	w	h	z	w	k	p	w	f	b	v
14.	v	v	b	b	f	f	h	h	f	h
15.	ʃ	p	f	f	p	h	v	v	p	f
16.	dʒ	f	p	j	b	b	p	p	h	b
17.	j	b	v	g	ʃ	ŋ	j	b	ŋ	g
18.	tʃ	ʃ	ŋ	h	g	ʃ	g	g	g	p
19.	h	ŋ	g	ʃ	ŋ	g	ŋ	ŋ	ʃ	ŋ
20.	z	g	ʃ	ŋ	j	j	ʃ	θ	j	θ
21.	ŋ	θ	θ	tʃ	θ	tʃ	θ	ʃ	dʒ	ʃ
22.	θ	dʒ	dʒ	dʒ	tʃ	dʒ	ʒ	tʃ	tʃ	tʃ
23.	θ	tʃ	tʃ	θ	dʒ	θ	ᴍ	dʒ	θ	dʒ
24.		ʒ	ʒ	ᴍ	ʒ	ʒ		ᴍ	ʒ	ᴍ
25.				ʒ				ʒ		ʒ

The consonant lists of the original studies
transformed to IPA and ranked according to frequency.

has reduced the number of consonants in each list to 22. Since the re-interpretation of unit phones as sequences increases the sample size, the whole system of frequencies has been altered. For this reason and because our calculations do not include vowel occurrences, the actual percentages used in our comparisons (Table 3) differ substantially from those presented by the original authors.

Correlation coefficients among the studies

Table 4 presents the results of computing the correlation coefficient *r* for each pair of lists. In attempting to interpret these results, the following factors should be considered: (1) the size of the sample, (2) the general nature and content of the sample, (3) the dialect from which the sample was collected, and (4) the method of transcription of the sample.

With respect to the first consideration, it appears that sample size, within the range represented in the ten studies, is not of great importance. This is evidenced, for

TABLE 3

	TRN	FOW	CAR	HAY	WHI	DEW	VOE	FRE	FRY	TOB
t	12.47	10.21	11.35	12.70	10.14	12.13	11.62	13.96	10.72	15.68
n	7.39	9.44	9.02	12.43	10.61	11.48	11.81	12.60	12.51	10.70
r	7.04	9.35	7.85	11.10	11.68	10.91	10.47	9.86	4.80	5.51
s	9.23	6.05	5.58	7.65	7.36	7.22	7.52	5.41	7.94	6.22
d	7.13	6.83	4.97	5.80	8.49	7.53	8.24	6.67	9.06	6.76
l	10.00	5.71	4.64	5.71	6.03	5.93	6.29	6.82	6.04	6.02
ð	0.41	4.37	3.61	5.24	6.01	5.44	5.10	5.54	5.87	6.02
k	8.22	4.43	3.80	4.66	3.41	4.30	4.13	4.57	5.10	4.88
m	5.16	3.83	4.62	4.49	4.80	4.41	5.05	4.95	5.31	5.26
w	2.21	6.41	8.39	3.35	3.63	3.30	2.97	4.55	4.64	4.93
j	1.69	8.65	12.36	1.88	1.04	0.95	1.87	2.95	1.45	3.99
z	1.34	4.12	3.14	3.69	4.58	4.71	3.45	3.05	4.06	3.54
v	1.93	2.97	1.83	3.64	3.72	3.61	2.48	2.41	3.30	2.93
h	1.55	3.61	5.59	2.31	3.67	2.87	3.24	2.87	2.41	2.85
f	2.63	2.61	2.10	2.52	3.23	2.92	3.47	2.57	2.95	2.46
b	4.88	2.22	2.67	2.58	2.58	2.87	3.16	2.28	3.25	2.40
g	3.22	1.03	1.55	1.78	1.24	1.17	1.74	2.03	1.73	2.31
p	6.13	2.62	2.05	3.51	2.68	3.24	2.39	2.33	2.94	2.28
ŋ	1.32	1.30	1.61	1.25	1.24	1.52	1.66	1.90	1.90	2.11
ʃ	3.39	2.37	1.51	2.19	2.18	2.12	1.63	1.36	2.26	1.33
θ	0.76	0.96	0.96	0.69	0.91	0.59	1.05	0.88	0.61	1.35
ʒ	1.90	0.91	0.80	0.83	0.77	0.78	0.66	0.44	1.15	0.47

Relative frequency of English
consonants as normalized in the present study.

TABLE 4

	HAY	VOE	WHI	FRE	TOB	FRY	FOW	TRN	CAR
DEW	985	984	982	956	871	897	810	735	580
HAY		972	952	966	884	879	828	748	626
VOE			975	963	879	909	825	751	617
WHI				918	806	863	801	667	558
FRE					945	891	866	726	712
TOB						898	821	714	725
FRY							737	721	545
FOW								584	922
TRN									415

Correlation matrix for the ten studies.

example, by the high degree of correlation between Whitney (6000 consonants) and Voelker (400,000 consonants).

Some of the major differences in the general nature and content of the samples are indicated in Table 1 under " Sample source ". Only the studies by Voelker (1937) and French *et al.* (1930) are based on samples taken from conversational materials. Both of these are in the group showing highest correlations, which indicates relatively high comparability between conversational and non-conversational materials. Also within this group are studies based on the very formal style of literary prose (Whitney, 1874) and on colloquial telephone conversations (French). Fowler (1957) describes this observation in the following way: " The distributional arrangements of phonemes is a part of the structure of the language which is not significantly disturbed by any individual differences either in author or subject."

On the other hand, the distinction between sampling from a dictionary and sampling from continuous texts is reflected in different results, as shown by the low correlation figures for Trnka's (1935) study. Also, the proportion of consonants to vowels is higher in the Trnka study than in any of the others. It is interesting that the correlation coefficients for Trnka's study, with the exception of those with Fowler and Carroll (to be discussed below) are all of about the same value. A great deal of the difference between Trnka's study and the rest can be attributed to a small number of phone types. Most markedly different is /ð/. In connected text materials, this phone has a relatively high frequency of occurrence, being the 6th most frequent consonant in four of the studies. In Trnka's study, it is the least frequent consonant. The obvious explanation is that while this phone occurs in only a small number of words, these words are of very frequent occurrence.

The fourth factor, difference in method of transcription, accounts for the great divergence shown in the study by Carroll (1952), the fairly large divergence of Fowler (1957), and to some extent the intermediate position of Tobias (1959).

The studies by Carroll and Fowler show a high degree of correlation, but a relatively low correlation with the other studies. This may be explained by the transcription method which was used, which is essentially that proposed by Trager and Smith (1951). In this notation the syllable nuclei of words such as *beat, bait, bite* end with /j/ (their /y/), and the syllable nuclei of words such as *boot, boat, bout* end with /w/. Furthermore, where certain syllabic nuclei are either lengthened or glided centrally, the symbol /h/ is employed (extensively by Carroll, much less by Fowler) as the second member of the nucleus. Consequently the values given for /j/, /w/, and /h/ by Carroll and Fowler have a rather wide divergence in Table 3 from those given in the other studies. Also Table 1 shows that the percentages of consonants for these two studies are markedly higher than for the others. The difference between the study by Fowler and that by Carroll is at least partly due to the way in which the syllabic /r/ is interpreted, e.g., *fur,* Fowler /fr/, Carroll /fir/.

The analysis by Tobias is based on the sample collected by French, and there is a high correlation between these studies. However, Tobias used the transcription proposed by Kenyon and Knott, in which the syllable nuclei in words like *bird* (as pronounced in the Midwestern United States) is represented by a single vowel symbol. Table 3 shows the very low percentage of /r/ in the study of Tobias as compared with that of the other studies.

Conclusion

In this study a comparison has been made of the consonant data from ten different frequency counts of English phonemes. The following major conclusions are based on an analysis of the data.

1. The relative frequency of consonants is not materially affected by the style or literary content of the sample. Also, the relative frequency data are rather stable with a sample size as small as 6000 items (Whitney).

2. Except for the obvious instances of /r/, the consonant frequencies do not vary significantly among the dialects which were considered here. This is, of course, a property of English, where dialectal differences are manifested mostly in the vowels. There are numerous languages which exhibit dialect differences primarily in the consonant system, or in prosodic features.

3. Dictionary sampling in English yields statistical results which are very different from those obtained by sampling continuous speech. In some languages this may not be the case. For example, Fant and Richter (1958) state that for Swedish " a phoneme frequency count based on every word counted once only leads to a distribution similar to the one found when every word has been weighted with its appropriate frequency-of-occurrence in the material."

4. Transcriptional differences may yield results which are not inter-convertible and consequently which are not comparable.

5. It is seen from Table 4 that the group of frequency counts reviewed here have good agreement for consonants. For application in mechanical speech recognition, it seems likely that higher order probabilities may be more valuable. In an actual recognizer, the use of statistical information should be dependent upon the certainty of the acoustical information (Peterson, 1959). The collection of such statistical information for application to automatic speech recognition can be fruitful only if the factors discussed above are properly considered in the collection of the data.

References

Fant, C. G. M. and Richter, M. (1958). Some notes on the relative occurrence of letters, phonemes, and words in Swedish. *Proceedings of the Eighth International Congress of Linguists* (Oslo), 815.

Fowler, M. (1957). Herdan's statistical parameter and the frequency of English phonemes. *Studies Presented to Joshua Whatmough* (The Hague), 45.

Fry, D. B. and Denes, P. (1957). On presenting the output of a mechanical speech recognizer. *J. acoust. Soc. Amer.*, **29**, 364.

Fry, D. B. and Denes, P. (1958). The solution of some fundamental problems in mechanical speech recognition. *Language and Speech*, **1**, 35.

Hoel, P. G. (1954). Introduction to Mathematical Statistics (New York).

Kenyon, J. S. and Knott, T. A. (1944). A Pronouncing Dictionary of American English (Springfield, Mass.).

Peterson, G. E. (1959). Linguistic concepts in automatic speech recognition procedures. *Proceedings of the Seminar on Speech Compression and Processing* (AFCRC-TR-59-198 Bedford, Massachusetts), 2.

Trager, G. L. and Smith, H. L. (1951). An outline of English structure. *Studies in Linguistics*, Occasional Papers No. 3.

Appendix

Frequency Studies of English Phonemes (see Tables).

TRN Trnka, B. (1935). A phonological analysis of present-day standard English. Studies in English by Members of the English Seminar of the Charles University (Prague), Vol. 5.

FOW Fowler, M. (1957). Herdan's statistical parameter and the frequency of English phonemes. *Studies Presented to Joshua Whatmough* (The Hague), 45.

CAR Carroll, J. B. (1952). Progress report on Project 52, Transitional Probabilities of English Phonemes, March 15, 1952; supplemental information of Project 52, October 27, 1952; hectograph materials privately distributed.

HAY Hayden, R. E. (1950). The relative frequency of phonemes in general American English. *Word*, **6**, 217.

WHI Whitney, W. D. (1874). The proportional elements of English utterance. *Proc. Amer. Philol. Assn.*, 14.

DEW Dewey, G. (1923). The Relative Frequency of English Speech Sounds (Cambridge).

VOE Voelker, C. H. (1937). A comparative study of investigations of phonetic dispersion in connected American speech. *Arch. néerl. de Phon. expér.*, **13**, 138.

FRE French, N. R., Carter, C. W., Jr. and Koenig, W. (1930). The words and sounds of telephone conversations. *Bell Telephone Syst. Tech. Publ.*, Monograph B-491.

FRY Fry, D. B. (1947). The frequency of occurrence of speech sounds in Southern English. *Arch. néerl. de Phon. expér.*, **20**, 103.

TOB Tobias, J. V. (1959). Relative occurrence of phonemes in American English. *J. acoust. Soc. Amer.*, **31**, 631.

RECURRENTLY IMPULSED RESONATORS IN SPEECH AND PSYCHOPHYSICAL STUDIES*

OM P. GANDHI, GORDON E. PETERSON AND FRANCIS YU
Speech Research Laboratory, University of Michigan

The need for data on the perception of sounds produced by the excitation of resonance systems is discussed in relation to speech and psycho-acoustics. The voltage response of a series of N decoupled low-pass resonator sections to the sudden application of various types of input pulse trains is reviewed.

The output of such a series of resonators to a recurrent impulse which is suddenly applied contains a d.c. term, a series of transient terms at the uncoupled resonator frequencies, and a steady-state term involving the harmonics of the input pulse train. A circuit for psychophysical tests is described which provides a number of discrete positions of resonator frequency and damping. Sound spectrographic analyses demonstrating the transient and steady-state terms for a single resonator are presented.

INTRODUCTION

It is well known that the laryngeal input wave to the vocal tract does not have an ideal pulse form (Steinberg, 1934). While many have studied the nature of the glottal wave, its characteristic shape is still somewhat obscure. This wave shape is primarily associated with certain types of vocal qualities. In general, in the simpler laryngeal qualities, the wave appears to be a monotonic quasi-periodic function whose spectral components decrease in amplitude as a function of frequency.

The development of a transmission line theory and analogue of the vocal tract provided a major advance in understanding the acoustical characteristics of the system to which the laryngeal input is applied (Dunn, 1950). More recent contributions to vowel theory have provided an extension of the transmission line analogue to include dissipation and lumped constant simplifications (Stevens, Kasowski, and Fant, 1953).

PSYCHOPHYSICS OF SPEECH

Although much is known about human responses to simple sounds, thus far there has been little success in relating psychophysical knowledge to an understanding of speech

* *This research was supported by the Information Systems Branch of the Office of Naval Research of the United States Navy, under Contract Nonr 1224(22), NR 049-122. This paper also appears as part of Report No. 6 of the Speech Research Laboratory of the University of Michigan:* " Studies on Impulsed Resonators ", *C. D. Creelman, O. P. Gandhi, G. E. Peterson, and F. Yu, February, 1960.*

perception. The psycho-acoustical literature contains many reports on the response of the human to pure tones, clicks, and random noise. While the sources involved in generating the various speech sounds have the properties of pure tone complexes, clicks, and random noise, these sounds are normally modified in speech production by the various resonating chambers of the vocal cavities. Thus there is actually very little psychophysical information available on responses to elementary sounds of the type generated in speech, and at present we have little basis for relating psychophysical measurements to data on speech.

There is considerable reason to believe that the human observer interprets meaningful sounds in terms of the various properties of the source, rather than according to the acoustical dimensions and magnitudes of the sounds or according to psycho-acoustical abstractions such as auditory scales. Thus psychophysical data, based on abstract sounds, will never provide a full understanding of speech perception. For example, vowels which are weaker in all essential amplitude properties may be judged as more stressed or louder than vowels of greater amplitude (Lehiste and Peterson, 1959). While a similar phenomenon has not yet been demonstrated for judgments of the pitch of vowels, it seems very probable that measurements of the fundamental frequency and harmonic amplitudes of vowels will not provide the information necessary to predict judgments of their pitch consistently. At least in the case of vowel amplitude judgments, a more coherent explanation lies in the properties of the speech source. The application of the more rigorous techniques which have been developed in psychophysics (Licklider, 1959) to the study of speech parameters should be of considerable value in achieving an increased understanding of speech perception.

Impulsed resonators

Toward the above objective a study has been undertaken to apply the techniques of signal detection and recognition research to the perception of elementary vowel formant patterns. The initial experiment involved a single resonator which was excited by a recurrent impulse.

The primary objective of the present paper is to discuss the response of one or more decoupled resonators to a recurrent input pulse train. It has long been known that if a sinusoid is suddenly applied to a single resonant system, the output time function will consist of a sinusoid of the same frequency plus a transient term corresponding to the natural frequency of the resonator. If such an input is suddenly applied to a series of such resonators, each decoupled from the next as shown in Fig. 1, then the output will consist of the sinusoid plus a series of transient terms corresponding to the natural frequencies of each of the resonators.

Throughout the present paper emphasis will be placed upon the type of resonant system shown in Fig. 1 because of its relation to speech production. If it is assumed that at audio frequencies no zeros occur in the transfer function of the vocal tract during vowel production, then vowel formations may be approximated by such a series of decoupled resonators.

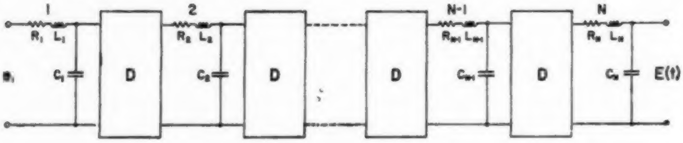


Fig. 1. An N-section low-pass resonator to which the input illustrated in Fig. 2 is applied as e_1. D represents decoupling networks of unity transfer function.

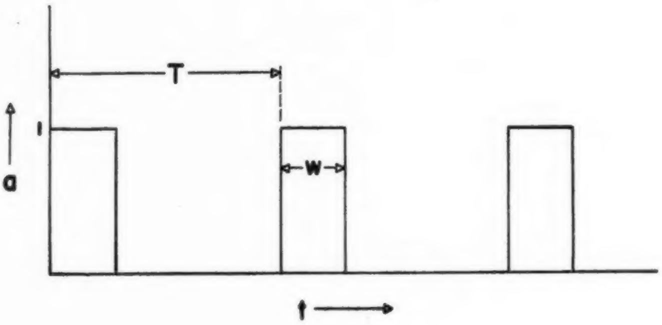


Fig. 2. Input function of recurring pulses of unit amplitude, width w, and period T.

Various authors (Flanagan, 1957, and Weibel, 1955) have described the response of an electrical speech analogue containing a series of poles to a single impulse. If a train of sharp pulses is suddenly applied to a series of decoupled resonators of the form shown in Fig. 1, the resulting output will in general contain three types of terms.

(a) If the pulses are of a single polarity, then the input wave contains a d.c. term, and the output will also contain a d.c. term. If the transfer function of the resonator system has band-pass or high-pass resonator characteristics, then no such term will appear at the output.

(b) A series of decaying transient terms will appear at the output of such an impulsed resonator corresponding to the natural frequency and damping characteristics of each of the resonators.

(c) Third, there will be a series of terms which represent the harmonics (Fourier series a.c. components) of the input pulse train.

RECTANGULAR INPUT PULSES

The same three types of terms appear in the time function of the output of a resonator if a train of rectangular pulses is suddenly applied. The derivation of the expression for the output in this case is somewhat more complicated than than for the sinusoidai or ideal pulse train case discussed above. Since the sudden application of an excitation function to a resonant system is of considerable interest in speech and psychophysics, the derivation for the case of rectangular pulses will be given in detail (van Valkenburg, 1955).

An input of recurring pulses of unit amplitude, width w, and period T is illustrated in Fig. 2. These can be expressed in Laplace transform notation as follows:

$$L[f(t)] = \int_{0+}^{\infty} e^{-St} f(t)dt$$

$$= \int_{0}^{w} e^{-St}\,dt + \int_{T}^{T+w} e^{-St}\,dt + \int_{2T}^{2T+w} e^{-St}\,dt + \cdots$$

$$= \left[1 + e^{-ST} + e^{-2ST} + \cdots + e^{-(n-1)ST}\right] \int_{0}^{w} e^{-St}\,dt$$

$$= \frac{1 - e^{-wS}}{S} \cdot \frac{1 - e^{-nST}}{1 - e^{-ST}} \tag{1}$$

where

n is the number of pulses which have occurred since the initiation of the pulses.

For a single impulse n = 1 and $(1\text{-}e^{-wS})/S$ is the difference between two step functions, one beginning at $t = 0$ and the other at $t = w$.

The transfer function of an RLC resonator can also be expressed in Laplace transform notation. If the input and output voltages are defined as e_i and e_o, then:

$$\frac{e_o(S)}{e_i(S)} = \frac{1/SC}{R + SL + 1/SC} = \frac{1}{S^2LC + RCS + 1} = \frac{S_a\,S_a^*}{(S-S_a)\,(S-S_a^*)} \tag{2}$$

where

$$S = (\sigma + j\omega)$$

$$\left.\begin{matrix} S_a \\ S_a^* \end{matrix}\right\} = -\frac{R}{2L} \pm j\sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}$$

$$= -\alpha \pm j\beta$$

$$\alpha = \frac{R}{2L} ; \qquad \beta = \left[\frac{1}{LC} - \frac{R^2}{4L^2}\right]^{1/2}$$

$$S_a S_a^* = \alpha^2 + \beta^2 = 1/LC$$

For N such decoupled resonators in sequence (Fig. 1) the total transfer function is:

$$\prod_k \frac{S_k S_k^*}{(S-S_k)\ (S-S_k^*)} \qquad (3)$$

where

$$k = 1, 2, \cdots N$$

$$S_k = -\alpha_k + j\beta_k$$

$$\alpha_k = + \frac{R_k}{2L_k} , \qquad \beta_k = \left[\frac{1}{L_k C_k} - \frac{{R_k}^2}{4{L_k}^2}\right]^{1/2}$$

If such a network of resonators has the initial condition of zero energy, then a series of recurrent pulses applied to the input will produce the following output:

$$e_o(S) = \frac{1 - e^{-wS}}{S} \times \frac{1 - e^{-nST}}{1 - e^{-ST}} \prod_k \frac{S_k S_k^*}{(S-S_k)\ (S-S_k^*)}$$

As n increases the term e^{-nst} rapidly becomes negligible, so that we may write:

$$e_o(S) = \frac{1 - e^{-wS}}{S(1-e^{-ST})} \prod_k \frac{S_k S_k^*}{(S-S_k)\ (S-S_k^*)} \qquad (4)$$

The simple poles of such a function in the complex S plane are shown in Fig. 3. The pole at S = 0 represents the d.c. term in the output and the N conjugate pairs of poles in the left half of the plane correspond to S_1, S_1^*, S_2, S_2^*, ... S_n, S_n^*. Each of these pairs of poles is represented by exponentially decreasing sinusoids at the output.

The output voltage $e_o(t)$ is found by taking the inverse Laplace transformation of Eq (4). $e_o(t)$ (indicated in Fig. 1) will be the sum of the residues of $e_o(S)$ at its poles.

Let

$$\prod_k S_k S_k^* \triangleq A = \frac{1}{L_1 C_1 L_2 C_2 \cdots L_n C_n} = \prod_k (B_k)$$

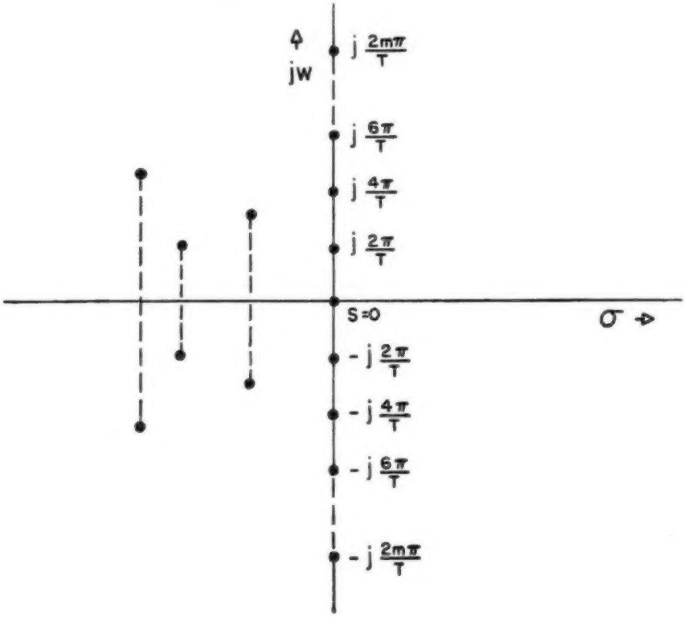


Fig. 3. Representation of the poles of Eq. (4) in the complex S plane.

where

$$B_k = \frac{1}{C_k L_k} \qquad K = 1, 2, \cdots N$$

Then the sum of the residues at $S = S_i$ and $S = S_i{}^*$ will be given by:

$$\left[e^{S_i t} \lim_{S \to S_i} \frac{1 - e^{-wS}}{(1-e^{-ST})S} \; \frac{A}{(S-S_i^*)} + e^{S_i^* t} \lim_{S \to S_i^*} \frac{1 - e^{-wS}}{(1-e^{-ST})S} \; \frac{A}{(S-S_i)} \right] \prod_{k \neq i} {}_k \frac{1}{(S-S_k)\;(S-S_k^*)} \; .$$

$$= \frac{A\ell_i \; e^{-\alpha_i t}}{\beta_i B_i \left\{ \prod_{\substack{k=1,2\cdots N \\ i \neq k}} {}_k \; M_{ik} \right\}} \; \text{Sin} \; (\beta_i t + \Phi_i) \qquad (5)$$

where

$$\ell_i \triangleq \left\{\frac{1 + e^{2w\alpha_i} - 2e^{w\alpha_i} \cos w\, \beta_i}{1 - 2e^{\alpha_i T} \cos(\beta_i T) + e^{2\alpha_i T}}\right\}^{1/2}$$

$$\Phi_i \triangleq \emptyset_i + \Theta_i - \left(\sum_{\substack{k=1\\k\neq i}} \psi_{ik} - \pi\right)$$

$$\emptyset_i \triangleq \tan^{-1}\left[\frac{\sin w\, \beta_i}{e^{-w\alpha_i} - \cos w\, \beta_i}\right] - \tan^{-1}\left[\frac{\sin \beta_i T}{e^{-\alpha_i T} - \cos \beta_i T}\right]$$

$$\psi_{ik} \triangleq \tan^{-1}\left\{\frac{2\beta_i(\alpha_k-\alpha_i)}{(\alpha_i-\alpha_k)^2 + (\beta_k{}^2-\beta_i{}^2)}\right\}$$

and

$$\Theta_i = \tan^{-1}\frac{\beta_i}{\alpha_i}$$

$$B_i = \sqrt{\alpha_i{}^2 + \beta_i{}^2}$$

$$M_{ik} = \left\{[(\alpha_i-\alpha_k)^2 + \beta_i{}^2 + \beta_k{}^2]^2 - 4\beta_i{}^2\beta_k{}^2\right\}^{1/2}$$

The total contribution to the output due to these conjugate pairs of poles in the left half plane is:

$$A \sum_{i=1}^{n} \frac{\ell_i\, e^{+\alpha_i t}}{\beta_i B_i \left\{\prod_{k\neq i} M_{ik}\right\}} \sin(\beta_i t + \Phi_i) \qquad (6)$$

The residue at the pole corresponding to S = 0 represents the d.c. output:

$$\lim_{S\to 0} A \left(\frac{1-e^{-wS}}{1-e^{-ST}}\right) \prod_k \frac{1}{(S-S_k)\,(S-S_k^*)} = \frac{w}{T} \qquad (7)$$

The sum of the residues (see Fig. 3) at $S = \pm j(2\pi m/T)$ is given by the expression:

$$A\left[e^{j(2\pi mt/T)} \lim_{S\to j(2\pi m/T)} \left\{\frac{[S-j(2\pi m/T)]\ (1-e^{-wS})}{(1-e^{-ST})S} \prod_k \frac{1}{(S-S_k)\ (S-S_k^*)}\right\}\right.$$

$$\left. + e^{-j(2\pi mt/T)} \lim_{S\to j(2\pi m/T)} \left\{\frac{[S+j(2\pi m/T)]\ (1-e^{-wS})}{(1-e^{-ST})S} \prod_k \frac{1}{(S-S_k)\ (S-S_k^*)}\right\}\right]$$

$$= A\left[\frac{R_m}{\pi m} \prod_k \frac{1}{S_{km}}\right] \sin\left(\frac{2\pi mt}{T} - \gamma_m - \sum_k \delta_{km}\right) \qquad (8)$$

where

$$R_m = \left[2 - 2\cos\frac{2\pi mw}{T}\right]^{1/2}$$

$$S_{km} = \left[\left(\alpha_k{}^2 + \beta_k{}^2 - \frac{4\pi^2m^2}{T^2}\right)^2 + \left(\frac{4\pi m\alpha_k}{T}\right)^2\right]^{1/2}$$

$$= \left[\left(B_k{}^2 - \frac{4\pi^2m^2}{T^2}\right)^2 + \left(\frac{4\pi m\alpha_k}{T}\right)^2\right]^{1/2}$$

$$\gamma_m = \tan^{-1}\left\{\frac{\sin\ (2\pi mw/T)}{1 - \cos\ (2\pi mw/T)}\right\}$$

$$\delta_{km} = \tan^{-1}\left\{\frac{4\pi m\alpha_k}{T\ [B_k{}^2 - (4\pi^2m^2/T^2)]}\right\}$$

Thus the output $e_o(t)$ of the resonator is given by:

$$e_o(t) = \frac{w}{T} + A \sum_{i=1}^{n} \left[\frac{\ell_i \, e^{-\alpha_i t}}{\beta_i B_i \left\{ \prod_{k \neq i} M_{ik} \right\}} \sin(\beta_i t + \Phi_i) \right.$$

$$\left. + \sum_{m=1}^{\infty} \frac{R_m}{\pi m \left\{ \prod_{k \neq i} S_{km} \right\}} \sin\left(\frac{2\pi m t}{T} - \gamma_m - \sum_k \delta_{km} \right) \right] \qquad (9)$$

The terms of (9) include:

(a) A d.c. term of value w/T.

(b) A series of transient terms of exponentially decreasing sinusoids of each of the resonance circuits of frequencies $\beta_1, \beta_2 \ldots \beta_n$ and of logarithmic decrements $\alpha_1, \alpha_2, \ldots \alpha_n$.

(c) The harmonics (including the fundamental) of the input frequency 1/T.

Single resonator

For the special case of recurrent periodic pulses applied to a single resonator:

$$e_o(t) = \frac{w}{T} + \frac{1}{L_1 C_1} \left[\frac{\ell_1 \, e^{-\alpha_1 t}}{\beta_1 B_1} \sin(\beta_1 t + \Phi_1) \right.$$

$$\left. + \sum_{m=1}^{\infty} \frac{R_m}{\pi m S_{1m}} \sin\left(\frac{2\pi m t}{T} + \gamma_m - \delta_{1m} \right) \right] \qquad (10)$$

where

$$\ell_1 = \left[\frac{1 + e^{2w\alpha_1} - 2e^{w\alpha_1} \cos w\beta_1}{1 - 2e^{\alpha_1 T} + e^{2\alpha_1 T}} \right]^{-1/2}$$

$$\Phi_1 = \tan^{-1} \left\{ \frac{\sin w\beta_1}{e^{-w\alpha_1} - \cos w\beta_1} \right\} - \tan^{-1} \left\{ \frac{\sin \beta_1 T}{e^{-\alpha_1 T} - \cos \beta_1 T} \right\}$$

$$+ \tan^{-1} \frac{\beta_1}{\alpha_1} - \pi$$

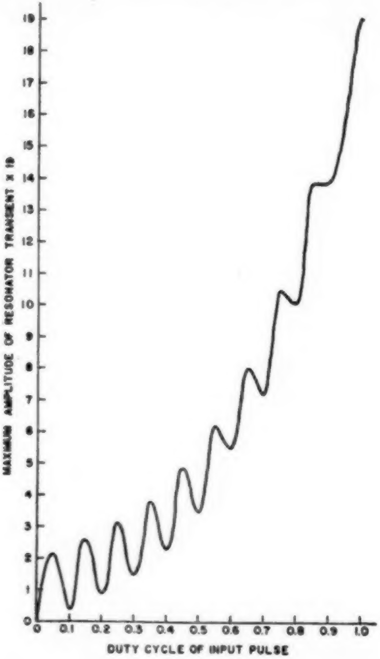


Fig. 4. The maximum amplitude of the resonatory transient [second term in Eq. (9)] as a function of input pulse width in milliseconds. The curve is illustrative; the values employed in calculating this specific curve were: $\alpha = 300 \times 10^{-1}$ seconds, $\beta = 2\pi \times 10^{-3}$ seconds, $T = 10^{-3}$ seconds.

$$\beta_1 = \frac{1}{\sqrt{L_1 C_1}} \quad ; \qquad \delta_{1m} = \tan^{-1}\left\{\frac{4\pi m \alpha_1}{T\ \beta_1^2 - (4\pi^2 m^2/T^2)}\right\}$$

$$S_{1m} = \left[\left(B_1^2 - \frac{4\pi^2 m^2}{T^2}\right)^2 + \left(\frac{4\pi m \alpha_1}{T}\right)^2\right]^{1/2}$$

The resonator frequency appears only as a transient in the output whose amplitude is:

$$\frac{1}{\beta_1 \sqrt{L_1 C_1}} \left[\frac{1 + e^{2w\alpha_1} - 2e^{w\alpha_1} \cos w\beta_1}{1 - 2e^{\alpha_1 T} \cos \beta_1 T + e^{2\alpha_1 T}}\right]^{1/2} e^{-\alpha_1 t}$$

If w is varied for a given value of T, the above amplitude of the transient term assumes the values shown in Fig. 4.

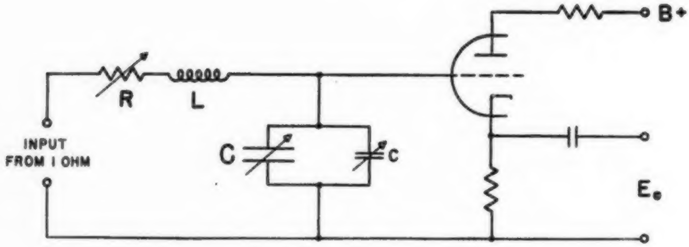


Fig. 5. Schematic diagram of resonator employed in psychophysical tests. The R, C, c indicated as variables actually involve a number of discrete positions.

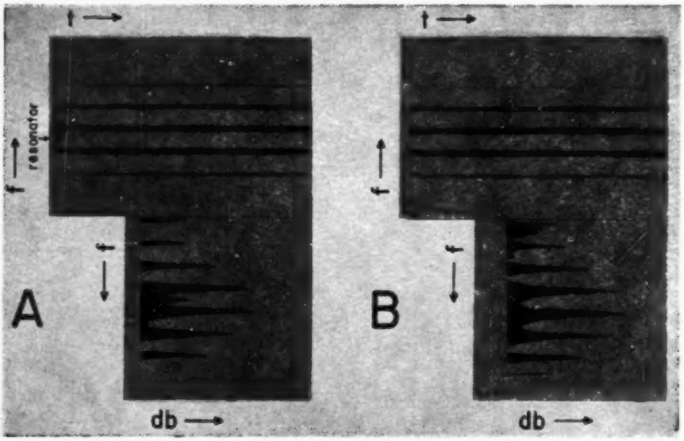


Fig. 6. Narrow filter sound spectrographic amplitude sections of the output of the recurrently impulsed resonator. (A) Analysis of initial period when pulses are applied to the resonator. (B) Analysis of resonatory output after steady-state conditions are reached.

For psychophysical studies (Tanner and Swets, 1954) a single resonator section has been constructed, as shown in Fig. 5, to which a recurrent impulse is applied. The various values of C provide seven resonance frequency range settings from 250 to 5350 cps. Within each range there are eleven incremental frequency settings (by means of c) varying from steps of 5 cps. for the lowest range to steps of 35 cps. for the highest range. Except as limited in the low range, for any given setting of C and c there are five different values of R, providing resonator bandwidths of approximately 5, 25, 100, 200, and 300 cps.

Illustrations of the second and third terms of Eq. (10) are given in the narrow filter spectrographic amplitude sections of Fig. 6. Both spectrograms are for the output of the same resonator, which has a natural resonance frequency of approximately 1700 cps. The input pulse rate was approximately 500 cps. In A of Fig. 6 the input pulse was suddenly applied to the resonator, and the transient term is clearly visible between the third and fourth harmonics. In B, however, the record was taken by suddenly closing a switch at the output of the resonator. In this latter spectrogram we do not find a transient term, since its amplitude had already decayed.

Exponentially decaying input pulses

Other general types of pulses which may be of interest are those of a saw-tooth or triangular form. Exponentially decaying pulses (as in the output of a relaxation oscillator) actually resemble the input laryngeal wave more closely than do pulses of an ideal or rectangular shape. Such pulses have been employed as an approximation to the laryngeal source (Dudley, Riesz, and Watkins, 1939).

In the case where the period T is much larger than the decay time constant of the pulses, the pulse amplitude $e^{-at} \rightarrow O$ within each cycle. Such a train of pulses of unit amplitude in Laplace transform notation is represented by

$$\frac{1}{(S+a)\ (1-e^{-st})}$$

where S is the Laplace variable.

When such a series of pulses forms the input to the resonator circuit of Fig. 1, the output is given by:

$$e_o(S) \;=\; \frac{1}{(S+a)\ (1-e^{-st})} \prod_k \frac{S_k\ S_k^*}{(S-S_k)\ (S-S_k^*)} \qquad (11)$$

The inverse Laplace transform of Eq. (11) gives the output of the resonators driven by recurring exponentially decaying unit pulses:

$$e_o(t) = \frac{A\,e^{-at}}{1 - e^{aT}} \prod_k \frac{1}{(a+\alpha_k)^2 + \beta_k^2}$$

$$+ \sum_{i=1}^{n} \frac{A\,\ell_i\,e^{-\alpha_i t}}{2\beta_i \prod_{k\neq 1} M_{ik}} \sin\left(\beta_i t + \Phi_i\right)\cdot$$

$$+ \sum_{m=1}^{\infty} \frac{2A\,R_m}{T \prod_{k\neq i} S_{km}} \cos\left(\frac{2\pi mt}{T} - \gamma_m + \sum \delta_{km}\right) \qquad (12)$$

where A, R_m, S_{km}, γ_m, S_{km}, M_{ik} are the same as defined previously.

$$\Phi_i = \phi_i + \theta_i + \sum \psi_{ik}$$

$$\theta_i = \tan^{-1}\left\{\frac{\beta_i}{-\alpha_i + a}\right\}$$

$$\phi_i = \tan^{-1} \left. \frac{e^{+\alpha_i T} \sin\ \beta_i T}{1 - e^{+\alpha_i T} \cos\ \beta_i T} \right\}$$

$$\ell_i = \frac{1}{\left[(1 - e^{+\alpha_i T} \cos\ \beta_i T)^2 + e^{+2\alpha_i T} \sin^2\ \beta_i T\right]^{1/2}}$$

Thus, except for magnitudes, the various terms correspond closely to those described in Eq. (9) for rectangular recurrent input impulse trains.

For very large relative values of T, the first and third terms in expression (12) become negligible and the output can be expressed by:

$$e_o(t) = \sum_{i=1}^{n} \frac{A\,\ell_i\,e^{-\alpha_i t}}{2\beta_i \prod_{k\neq i} M_{ik}} \sin\left(\beta_i t + \Phi_i\right) \qquad (13)$$

where

$$\Phi_i = \theta_i + \sum \psi_{ik} + \phi_i$$

This represents the sum of signals of frequencies β, having decaying amplitudes with decay constants α_i.

Table 1

SINUSOIDAL COMPLEX APPLIED TO INPUT	OUTPUT					
	Low Pass		Band Pass		High Pass	
	Initial	Terminal	Initial	Terminal	Initial	Terminal
d.c. present	i					
in input	ii	ii	ii	ii	ii	ii
wave	iii		iii		iii	
d.c. absent						
in input	ii	ii	ii	ii	ii	ii
wave	iii		iii		iii	

Terms which appear at the output of a series of isolated resonator sections when a periodic input is suddenly initiated or terminated.

i = d.c. term ;
ii = resonator transient terms ;
iii = harmonics of the input.

Discussion

The above review has been made of low-pass RLC resonator action, primarily because of its relevance to both psychophysical and phonetic problems. In general, three types of terms appear at the terminal output when a driving force which is periodic after the time t_0 is applied to a sequence of isolated low-pass resonators. For band-pass and high-pass resonator sections the d.c. component will be absent. The various cases are outlined in Table 1.

When the human vocal tract functions as a uni-dimensional acoustical tube of variable cross-sectional area in vowel production, it may be simulated by a series of low-pass isolated resonator sections such as those described above. When voicing is suddenly applied to the vocal tube or is suddenly interrupted, a series of oscillating transient terms will result. Similar transients may be expected when the fundamental voice frequency is varied. These terms may provide cues for vowel identification, of course, which are absent in steady-state sections of vowels. Various studies of sustained vowels (Peterson, 1954) show that these terms do not exert the primary control of vowel identification, but it seems possible that they may contribute to it, especially when the sustained part alone is for some reason obscure. As Eqs. (9) and (12) show, the sustained parts of the vowel waves actually do not contain the resonator frequencies. Thus in speech analysis, these frequencies can be recovered only by some non-linear or active analysis of the wave. Analysis of isolated voice periods (Peterson, 1959) and active inverse filtering (Miller, 1959) are two techniques which merit further examination.

References

Dudley, H., Riesz, R. R. and Watkins, S. S. A. (1939). A synthetic speaker. *J. Franklin Institute,* **227**, 739.

Dunn, H. K. (1950). The calculation of vowel resonances and an electrical vocal tract. *J. acoust. Soc. Amer.,* **22**, 740.

Flanagan, J. L. (1957). Note on the design of " terminal-analog " speech synthesizers. *J. acoust. Soc. Amer.,* **29**, 306.

Lehiste, I. and Peterson, G. E. (1959). Vowel amplitude and phonemic stress in American English. *J. acoust. Soc. Amer.,* **31**, 428.

Licklider, J. C. R. (1959). Three auditory theories. *Psychology : A Study of Science* (Study I. Conceptual and Systematic) ; Volume I : Sensory, Perceptual, and Psychological Formulations, edited by S. Kock (New York), 41.

Miller, R. L. (1959). Nature of the vocal cord wave. *J. acoust. Soc. Amer.,* **31**, 667.

Peterson, G. E. (1954). Systematic research in experimental phonetics : The evaluation of speech signals. *J. speech hear. Disorders,* **19**, 158.

Peterson, G. E. (1959). Vowel formant measurements. *J. speech hear. Research,* **2**, 173.

Steinberg, J. C. (1934). Application of sound measuring instruments to the study of phonetic problems. *J. acoust. Soc. Amer.,* **6**, 16.

Stevens, K. N., Kasowski, S. and Fant, C. G. M. (1953). An electrical analog of the vocal tract. *J. acoust. Soc. Amer.,* **25**, 734.

Tanner, W. P. and Swets, J. A. (1954). The human use of information : I. Signal detection for the case of the signal known exactly. *Transactions of the IRE Professional Group on Information Theory, PGIT-4,* 213.

Van Valkenburg, M. E. (1955). Network Analysis (New York), 153.

Weibel, E. S. (1955). Vowel synthesis by means of resonant circuits. *J. acoust. Soc. Amer.,* **27**, 858.

PERCEPTION OF CONSONANT VOICING IN NOISE*

J. M. PICKETT** AND HERBERT RUBENSTEIN
Operational Applications Office, Air Force Command and Control Development Division, Bedford, Massachusetts

Measurements are reported of the perception in noise of the occurrence of voicing in the English consonants /p, b, t, d, f, v, s, z/. The listeners' task was to report whether the consonant spoken was of the class /b, d, v, z/ (voiced) or of the class /p, t, f, s/ (unvoiced). The factors investigated were (1) the position of the consonant in the test utterance: initial, intervocalic or final; (2) the place of articulation: alveolar /t, d, s, z/, or labial, /p, f, b, v/; (3) the degree of occlusion: stop, /p, b, t, d/, or fricative, /f, v, s, z/, and (4) the spectrum of the masking noise: white noise or low-frequency noise. The absence of voicing was perceived better in alveolar consonants than in labials in low-frequency noise. Otherwise there were no large effects of position, place of articulation, or degree of occlusion, on voicing perception. The results are interpreted in terms of low-frequency cues to voicing which are independent of place of articulation and high-frequency cues which vary with place of articulation.

Perceptual dimensions of spoken consonants, as demonstrated by confusion tendencies, are closely related to dimensions of articulation (Miller and Nicely, 1955 ; Pickett, 1958). Furthermore, to a first approximation, the dimensions appear to operate independently of each other. The present paper is a study of the perceptual independence of one consonant dimension, the dimension of voicing.

Examination of Miller and Nicely's data on initial consonants indicated to us that the perceptual independence of voicing depended on the portion of the speech spectrum transmitted to the listener. When only the lower speech frequencies were heard (low-pass filtering or white noise masking), the perception of voicing appeared to be

* *This is Technical Note 59-10, AD 216767, of the U.S. Air Force Cambridge Research Center, now the AF Command and Control Development Division. This research was performed at the Operational Applications Laboratory as part of Project 7684, Communication in Noise, of the Air Research and Development Command's Noise and Vibration Programme. A preliminary report was presented at the Third International Congress on Acoustics, Stuttgart, 1959. The final analysis of the data, the analysis of Miller and Nicely's data, and the writing of this report were done at the U.S. Naval Research Laboratory. The authors are indebted to Dr. Irwin Pollack and John Schjelderup who designed and built the automatic matrix-recorder and to Cecelia Burke who analyzed the data.*

** *Now at U.S. Naval Research Laboratory, Washington, D.C.*

relatively independent of the other articulatory dimensions. When, on the other hand, the higher speech frequencies were heard (high-pass filtering), perception of voicing appeared to interact with both place and degree of occlusion ; specifically, voicing was perceived better in the case of alveolars than in labials and voicing of stops was perceived better than that of fricatives.

In the experiments reported below, a further look is taken at the perception of voicing—not only in the initial position studied by Miller and Nicely, but in the intervocalic and final positions as well. The perception of voicing is studied in combination with labial and alveolar place of articulation, and with fricative and stop manner of occlusion. All tests were carried out at various speech-to-noise (S/N) ratios in two noise spectra.

Methods

The consonants tested were /p, b, t, d, f, v, s/, and /z/. They were spoken one at a time in a random order and heard through noise by listeners who recorded whether the consonant spoken was one of the voiced group, /b, d, v, z/, or one of the unvoiced group, /p, t, f, s/. The percentage of correct assignments to the proper group was taken as the measure of perception of voicing.

Two male college students served as talkers. The listeners consisted of a crew of four college students who were phonetically naive but experienced in listening in noise. American English was the native language of all. The talkers were trained to talk at a constant average speech level. The vowel used for all test utterances was /ɪ/, as in the word *hit*. In a given testing run, the consonants were spoken in only one of the three positions: initial position, CV ; intervocalic position, VCV, or final position, VC. The talkers read intervocalic /t/ as a stop with release, rather than as an alveolar flap, a frequent American pronunciation. As a rule, the final consonants were spoken with releases (Malécot, 1958). Each test utterance was spoken fluently after the introductory carrier syllable, /tra/. The vowels were prolonged in utterance to about three times normal duration in order to reduce the influence of vowel duration as a cue to the voicing of the adjacent consonant (Denes, 1955).

Each testing run was made up of 20 utterances of each consonant. The series of consonants to be spoken was presented to the talker by an automatic matrix-recorder which also accumulated a matrix of frequencies of the stimulus-response combinations. A new stimulus consonant was presented every 2 to 3 seconds. Each listener was provided with a pair of response buttons ; one button was pushed for the unvoiced judgment, i.e., the class /t, p, s, f/, and the other button was pushed for the voiced judgment, the class /d, b, z, v/. Between utterances each listener's voicing judgment of the previous consonant was recorded by the matrix-recorder. The recorder accumulated on a set of 16 counters the listener's total number of correct and incorrect judgments separately for each of the 8 stimulus consonants.

The speech signal was received by a microphone and mixed electrically with the continuous noise signal. The mixed speech and noise were then amplified and presented to the listeners over earphones. Tests were carried out with two noise spectra: a white random noise having equal energy at all frequencies, and a low-frequency random noise. The spectrum slope of the low-frequency noise was – 12 db per octave over the range 250 - 6800 cycles ; it was flat below 250 cycles. The talker monitored his average speech level on a VU meter. The S/N ratios were based on the monitored level.[1]

For the final tests, after some initial practice, one run of 160 consonants was carried out for each talker under each combination of S/N ratio, noise spectrum, and consonant position in utterance. All the tests of final consonant position were carried out first for both noise spectra, then the tests of initial position, and finally the tests of intervocalic position.

Results

The listeners' individual responses, under a given combination of talker, position in utterance, noise spectrum, and S/N ratio, were pooled separately for the eight spoken consonants. In this form, the data were first examined for any gross differences between the two talkers. Both talkers showed the same general results although the effects referred to below were larger for one talker than the other. The talker showing the smaller effects seemed on casual observation to articulate rather more precisely than the average American college male. The final data were formed by combining the data of the talkers.

The results are shown in the graphs of Fig. 1. The ordinate of each graph is the percentage of correct voicing responses. Each abscissa is the S/N ratio in db. The graphs are arranged in pairs, one of a pair showing the results for stop consonants, /p, b, t, d/, and the other for fricative consonants, /f, v, s, z/. The pairs are arranged in three rows and two columns. The rows correspond to the three different consonant positions in utterance: initial, intervocalic, and final positions. The columns correspond to the two noise spectral conditions, white noise and low-frequency noise.

As a whole it is apparent from Fig. 1 that the perception of voicing is relatively independent of position in utterance and of the degree of occlusion (stops *vs.* fricatives). This may be seen by noting that there are no gross differences between either the rows of graphs (positions in utterance) or between the two members of any pair of graphs (degrees of occlusion), provided only the overall perception of voicing is

[1] *The microphone (RCA 88), mixer, and amplifier, had uniform frequency response over the range of the earphones (PDR-8). The noise spectra were measured at the input to the earphones and limited by a band-pass filter to the range 100-6800 cycles. S/N ratios were measured after the filter with a VU meter at the input to the earphones. The sound level of the noise under the earphones was 80 db re 0·0002 dynes/cm^2 for the white noise and 90 db for the low-frequency noise.*

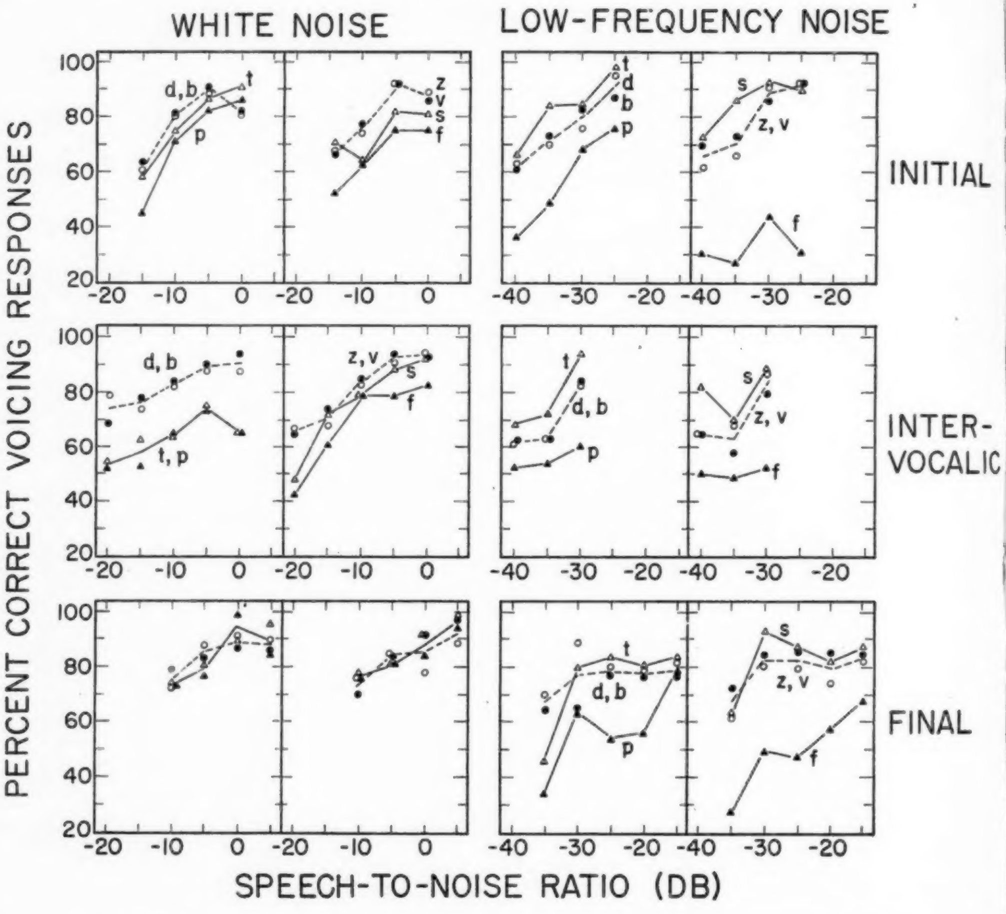


Fig. 1. The perception of consonant voicing in noise as related to noise spectrum, S/N ratio, and position in utterance. The ordinate of each graph is the percentage of correct voicing judgments. Each abscissa is the S/N ratio. The left column of graphs shows the results with a white noise spectrum; the right column of graphs is for a low-frequency noise spectrum. The three rows of graphs from top to bottom show results for initial, intervocalic, and final positions in the test utterances. Each point is based on 160 observations: 2 talkers each read 20 consonants to 4 listeners. The filled symbols indicate results with labial consonants, /p, b, f/ and /v/; the open symbols are for alveolar consonants, /t, d, s/ and /z/; the circles are for voiced consonants; the triangles are for unvoiced consonants. The main effects to be noted are the superior perception in low-frequency noise of the absence of voicing in alveolar unvoiced consonants, /t, s/, the poor perception of this with labial unvoiced consonants, /p, f/, and the absence of any large differences in voicing perception among all other conditions.

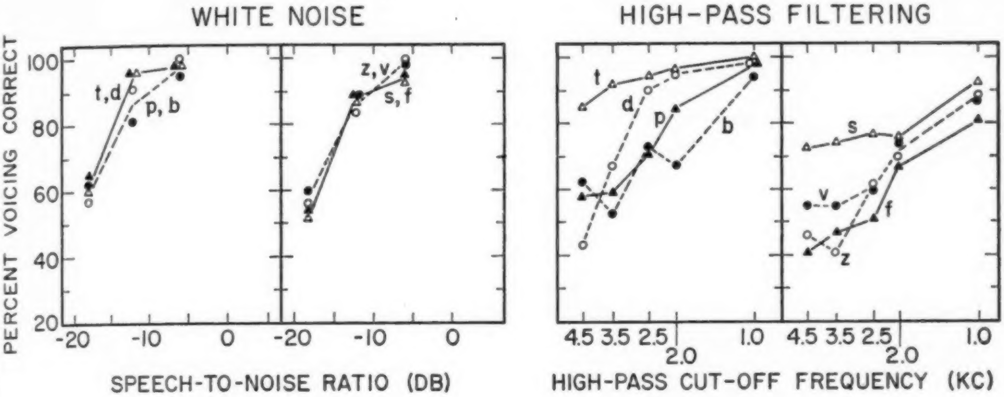


Fig. 2. Perception of consonant voicing in the experiment of Miller and Nicely (1955). The data are limited to conditions analogous to the top row of Fig. 1. The left pair of graphs shows voicing perception as a function of S/N ratio in a white noise (200 to 6500 cycles). The right pair of graphs shows voicing perception as a function of the speech band admitted between 5000 and 1000 cycles (white noise, speech level before filtering constant at S/N = 12 db, filter cut-off rate 24 db per octave). There were 103 to 273 observations per point. The points are coded as for Fig. 1. As in Fig. 1, the perception of absence of voicing is poor for the labial consonants /p, f/ when conditions are analogous to masking with low-frequency noise (high-pass filtering).

considered. A comparison of the two columns, however, reveals that the noise spectrum does have a very considerable effect. First, the dispersion of voicing perception as a function of the place of occlusion is very large in low-frequency noise relative to that in white noise. This will be seen to be due mainly to the unvoiced consonants in low-frequency noise where the absence of voicing in the alveolars, /t/ and /s/, was perceived much better than in the labials, /p/ and /f/. The voicing perception for the voiced consonants on the other hand, is independent of place of articulation (/d/ *vs.* /b/ and /z/ *vs.* /v/) in both noise spectra.

In white noise the voicing of the consonants /b, d, v, z/ (dashed lines) is perceived somewhat better than the absence of voicing of /p, t, f, s/ (solid lines). However, this effect is slight and occurs only in initial and intervocalic positions.

Discussion

As we have seen, place of articulation is of little significance in white noise but plays a very considerable role in low-frequency noise. The voicelessness of the alveolars /t, s/ is much more perceptible in low-frequency noise than the voicelessness of their labial counterparts /p, f/. Acoustic research on these consonants has established differences in the spectra of burst and fricative energy which lead to an explanation of this result. Halle *et al.* (1957) measured the spectra of the bursts following spoken /t/ and /p/. The bursts of /t/ contained energy between 4000 and 10,000 cycles while the bursts of /p/ were between 500 and 1500 cycles. In the perceptual experiments of Liberman *et al.* (1952), the frequency of the burst energy was varied to differentiate /t/ and /p/ in synthetic syllables. High-frequency bursts (mean of about 3600 cycles) led to judgments of /t/ and low-frequency bursts (mean of about 1800 cycles) led to judgments of /p/. The difference in high-frequency energy between /s/ and /f/ does not appear so clear cut. In the *normalized* spectra of Hughes and Halle (1956) the spectra of /s/ and /f/ are very similar. However, in view of the lower average power of /f/ (Fletcher, 1953, Table 6), it is apparent that, on an absolute basis, /s/ must have more energy than /f/ in the higher frequencies. The results of Hughes and Halle's perceptual experiment support this. They found the percentage of correct identifications to be highest when the maximum energy was located below 1500 cps. for /f/ and above 4000 cps. for /s/.

Now if we assume that, to a large extent, the low-frequency noise masks the voicing of a consonant and the vowel-formant cues to voicing, the only remaining cues might be the presence or absence of a phase of mid- or high-frequency consonant energy, the burst, aspirant or fricative phases associated with unvoiced consonants. Low-frequency noise, since it would mask the middle consonant frequencies more than the high frequencies, would accordingly leave the voicelessness of alveolars more perceptible than the voicelessness of the labials.

A corollary of this view of voicing perception in low-frequency noise would be that, when little or no aspirant or fricative energy is heard, that is when /p, f, b, v, d/, or /z/ are heard at a low S/N ratio, a voiced judgment is given, resulting in correct responses for the voiced consonants but incorrect responses for /p/ and /f/. Our data are consistent with this idea. With low-frequency noise we note in Fig. 1 for initial and final positions at the lowest S/N ratios, that the percentages of correct response to /p/ and /f/ are considerably less than chance (50%), i.e., there is a preponderance of " voiced " judgments. In white noise, correct responses to /p/ and /f/ at low S/N ratios are not significantly lower than chance. Furthermore, correct responses to /d, b, z, v/ are always greater than chance at the low S/N ratios.

In addition to our own data, an experiment by Lotz, Abramson, *et al.* (1960) tends to confirm the importance of hearing aspiration in judging the voicing of initial English stops. In this experiment, the unaspirated voiceless stops which occur in an initial cluster after /s/ were presented without the /s/ to native American listeners ; the stops were to be judged as either voiceless (/p/, /t/ or /k/) or voiced (/b/, /d/ or /g/) In spite of the voicelessness of these " residual " stops, they were judged to be voiced about 94% of the time (*op. cit.*, Fig. 1, p. 73).

Similar reasoning may be followed concerning the cues to voicing under masking by white noise. Assume that aspirant and fricative cues are masked by the white noise. Then it would follow from our data that the low-frequency cues to voicing operate independently of place or degree of occlusion. To our present knowledge, this is the case. Hughes and Halle (1956) show that both /v/ and /z/ have a strong component below 700 cps., the voice component, that is not found in /f/ and /s/. Other low-frequency cues to voicing exist in the early onset of the first vowel formant after voiced stops (Liberman *et al.*, 1958) and, for intervocalic consonants, in the shorter intervocalic interval of a voiced occlusion (Lisker, 1957). Thus, we would speculate that, for our tests with white noise where the middle and high-frequency cues are difficult to hear, the listener must rely largely on cues in the low-frequency speech energy, and that the detectability of these cues is independent of place or degree of occlusion.

Additional data on the spectral differentiation of voicing cues are available in the study of Miller and Nicely (1955). This was a study of the perception of initial consonants spoken before the vowel /a/. Although the listeners responded to more than one consonant dimension, the data published by Miller and Nicely may be treated like those of the present study by collapsing responses to each consonant into just two categories, voiced and unvoiced. This we did, restricting our analysis to the stimulus-response set of our own study and covering noise and frequency conditions analogous to ours. The results are shown in Fig. 2. The left pair of graphs in Fig. 2 shows the perception of voicing in white noise at various S/N ratios. The right pair of graphs shows perception of voicing as the frequencies of speech between 5000 and 1000 cps. are progressively admitted to the signal heard by the listeners, a situation analogous to increasing S/N ratio over a low-frequency noise ; the speech outside these frequency limits was removed by filtering. As we proceed from left to right on the scale of high-pass cut-off, the listener first hears only the high speech frequencies and then lower frequencies are added progressively down to 1000 cycles. Under these conditions of high- and mid-frequency listening it will be noted that the perception of the voicing of the voiced consonants /b, v, d, z/, and of the labial consonants /p, f/, requires hearing of the mid-frequencies below 2500 cps. On the other hand, the lack of voicing of the unvoiced alveolars, /t, s/, is perceived even when only a very restricted high-frequency region is heard. The perception of the absence of voicing in the labials /p/ and /f/ was relatively better than in our own study. This may mean that the mid-frequency energy of /p, f/ was heard clearly when the filtering allowed, as opposed to our own study where it was always partially masked. The voicing of stops was heard better than that of fricatives for mid- and high-frequency listening. Otherwise the results of Miller and Nicely in Fig. 2 are quite similar to our own results in the initial position (Fig. 1, top row).

The fricative consonants studied here are generally spoken with higher articulatory force (mouth pressure) than their stop counterparts (Black, 1950 ; Stetson, 1951). Nevertheless, the perception of the voicing of the fricatives was not superior to that of the stops. Black (1950) has also shown that initial consonant position yields higher consonantal mouth pressures than intervocalic or final positions. However, perception of voicing appears to be about equally good for all positions. Apparently then, the average mouth pressure differences behind different types of consonant occlusion have little effect on the perception of voicing, despite the fact that unvoiced occlusions result in higher pressures than voiced occlusions[2].

Conclusion

The perception of the voicing of the consonants /p, b, t, d, f, v, s, z/ was measured in two noise spectra: white noise and low-frequency noise. For the conditions examined, the results are interpreted to lead to the following conclusions:

1) Perception of voicing in both noise spectra is about equally good in initial, intervocalic, and final positions in utterance.

2) Perception of voicing is about equally good for stops and fricatives.

3) In white noise, where low-frequency voicing cues are presumed to dominate, voicing perception is independent of place of articulation within the consonant set examined. It is suggested that this result occurs because the low-frequency cues are such as to operate independently of place of articulation.

4) In low-frequency noise, where high-frequency voicing cues are presumed to dominate, the perception of the absence of voicing in the alveolars /t, s/ is better than that of the labials /p, f/. This result is held to be due to the larger amounts of high-frequency energy associated with the alveolar consonants.

[2] *The voiced-voiceless distinction in English might better be called a lenis-fortis distinction as suggested by Jacobson, Fant, and Halle in* Preliminaries to Speech Analysis, *Tech. Rept. 13, Acoustics Lab., Massachusetts Institute of Technology, 1952. The pressure differences in force behind the lenis (voiced) consonants versus the fortis (unvoiced) consonants have been noted and studied by Hudgins and Stetson (1935), Black (1950), and Stetson (1951). The occurrence of larynx vibration (voicing) during a consonant occlusion probably requires a relatively low air pressure behind the occlusion. In the present discussion the duration of occlusion and the opening of the occlusion into the burst and vowel phases of the syllable are assumed to afford the predominant perceptual cues as to the amount of pressure behind the occlusions. For current acoustic-articulatory theories on voicing, one should consult Meyer-Eppler (1953), Fischer-Jørgensen (1954), and Fant (1957, especially pages 22-24, 28, and Fig. 6). It is also interesting to note that many observers believe that the voiced-voiceless distinction is easily perceived in whispered speech.*

References

Black, J. W. (1950). The pressure component in the production of consonants. *J. Speech and Hearing Disorders,* **15**, 207.

Denes, P. (1955). Effects of duration on the perception of voicing. *J. acoust. Soc. Amer.,* **27**, 761.

Fant, G. (1957). Modern instruments and methods for acoustic studies of speech. *Stockholm Royal Inst. Technology Rep.,* No. 8, 1957 *(USAF Cambridge Res. Center Tech. Rep. TN58-112).*

Fischer-Jørgensen, E. (1954). Acoustic analysis of stop consonants. *Misc. phonetica* (IPA), II, 42.

Fletcher, H. (1953). Speech and Hearing in Communication (New York).

Halle, M., Hughes, G. W. and Radley, J.-P. A. (1957). Acoustic properties of stop consonants. *J. acoust. Soc. Amer.,* **29**, 107.

Hudgins, C. V. and Stetson, R. H. (1935). Voicing of consonants by depression of the larynx. *Arch. néerl. Phon. expér.,* **11**, 1.

Hughes, G. W. and Halle, M. (1956). Spectral properties of fricative consonants. *J. acoust. Soc. Amer.,* **28**, 303.

Liberman, A. M., Delattre, P. C. and Cooper, F. S. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *Amer. J. Psychol.,* **65**, 497.

Liberman, A. M., Delattre, P. C. and Cooper, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech,* **1**, 153.

Lisker, L. (1957). Closure duration and the intervocalic voiced-voiceless distinction in English. *Language,* **33**, 42.

Lotz, J., Abramson, A. S., Gerstman, L. J., Ingemann, F. and Nemser, W. J. (1960). The perception of English stops by speakers of English, Hungarian and Thai: a tape-cutting experiment. *Language and Speech,* **3**, 71.

Malécot, A. (1958). The role of releases in the identification of released final stops. *Language,* **34**, 370.

Meyer-Eppler, W. (1953). Untersuchungen zur Schallstruktur der stimmhaften und stimmlosen Geräuschlaute. *Z. Phon.,* **7**, 89.

Miller, G. A. and Nicely, P. (1955). Analysis of perceptual confusions among some English consonants. *J. acoust. Soc. Amer.,* **27**, 338.

Pickett, J. M. (1958). Perception of compound consonants. *Language and Speech,* **1**, 288.

Stetson, R. H. (1951). Motor Phonetics (Amsterdam).

A MODEL FOR SPEECH UNIT DURATION*

WM. A. HARGREAVES
The Langley Porter Neuropsychiatric Institute, San Francisco

A version of the exponential decay process is advanced as a mathematical model for speech unit durations. Distributions of speech unit durations are reported from a variety of speech samples. It is argued that a simple exponential model provides a workable approximation to these distributions. A more exact model is also outlined which takes account of (1) the fact that extremely short speech units lasting only a small fraction of a second are relatively less likely to occur, and (2) the fact that there tend to be small shifts in expected speech unit duration within a particular speech sample.

In studies which measure temporal patterns of speech, the form of the distribution of the duration of speech units has usually been reported, but with only passing interest. It is typically said to be J-shaped, such that the shortest speeches are most numerous, the frequency of occurrence decreasing in a regular fashion with increasing duration. A number of writers[1] have suggested or implied that a certain random process might be a useful model to approximate such observed distributions. The present paper will state one form of such a model, and explore its adequacy in the light of available data on speech unit durations.

The model is a version of the exponential decay process. In its simplest form it assumes that for a person who is emitting speech at some point in time, there is some fixed probability that he will stop speaking in the next time interval, and that this probability is independent of how long he has been speaking. As an example, assume that when a person begins speaking, there is a 50% chance that he will lapse into silence again within one second. Suppose, however, that he does not, and that he is still emitting speech at the end of five seconds. An exponential decay hypothesis would predict that there is then the same 50% chance that he will stop between five and six seconds. If we make a second assumption, that this same probability value of 0·5 applies during all the utterances in a particular sample, then the frequency distribution of these speech units should have the familiar exponential form. Half of them would end within the first second ; another quarter in the second second, and so on.

If we plot log frequency instead of raw frequency against duration, an exponential curve will form a straight line like the one at the top of Fig. 1. The slope of a particular curve corresponds to the probability of stopping. This is in turn the reciprocal of the mean unit duration in the population of utterances from which the

* *This study was supported in part by grants M-2015 and MY-3375, U.S. Public Health Service.*

[1] *Notably F. Mosteller in* "A model for speech and silence distributions," *an unpublished memorandum on the Verzeano-Finesinger analyzer (Verzeano & Finesinger, 1949).*

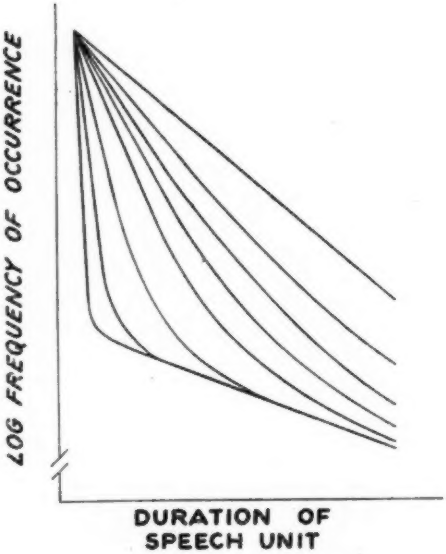


Fig. 1. Examples of the sum of two simple exponential distributions.

sample was drawn. By plotting experimental data on a log scale like this, one can get a visual impression of how well the data fit a simple exponential function, since lack of fit will be evident in the non-linearity of the curve.

I will argue here that this simple version of the model is adequate for a gross approximation to speech unit durations, but that the approximation can be improved by two elaborations of the model. In the first place, speech samples are rarely collected in situations where our second assumption is likely to hold. This is the assumption that the probability of stopping remains stationary over all the speech units in the sample. The mean duration in a sample reflects this probability value, and it is known that different persons have different tendencies with regard to mean speech unit durations (Matarazzo, Saslow, Matarazzo and Phillips, 1958). It is also known that mean duration changes for the same individual in different situations (Hargreaves, 1959) and will even change while he converses with the same person, if that person changes his speaking pattern (Matarazzo, *et al.*, 1958). Because of these facts, it will be useful to note the form of exponential distributions when the expected duration is not stationary. Fig. 1 shows examples of the simplest kind of combined pattern which might be produced by shifts in expected duration. These curves represent the situation where the expected duration takes on two values. They were generated by taking a sub-group with a long mean duration, corresponding to the relatively flat curve at the bottom, and adding to this a second group, whose mean was varied. When this mean is very

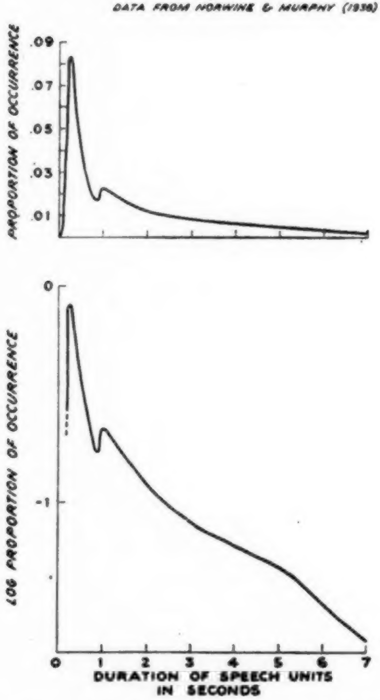

DATA FROM NORWINE & MURPHY (1938)

Fig. 2. Distribution of speech unit durations in long distance telephone conversations.

short, the composite curve shows a distinct inflection. When the mean is longer the curve takes on a scalloped appearance, and finally becomes indistinguishable from a single exponential. Consider now a composite sample composed of not just two, but multiple sub-samples with different means. In this case the result is typically a scalloped curve somewhat like the upper ones in the figure. This scalloped curve would also result if the probability of stopping drifted about during a particular sample. Thus the scalloped curve might be expected to appear quite frequently in samples of speech.

There is a second modification which is necessary before we can hope for the model to be generally applicable. We must take account of the fact that extremely short speech units are relatively rare. This can be illustrated with some data from telephone conversations shown in Fig. 2. The sample was collected in 1938 at the Bell Telephone Laboratories (Norwine and Murphy, 1938). The mode falls at a quarter second, which represents a monosyllabic comment or grunt, and there are very few units which are shorter than this modal value. We shall refer to this lack of very short units as the " start effect ". When data are grouped into one-second intervals, the existence of

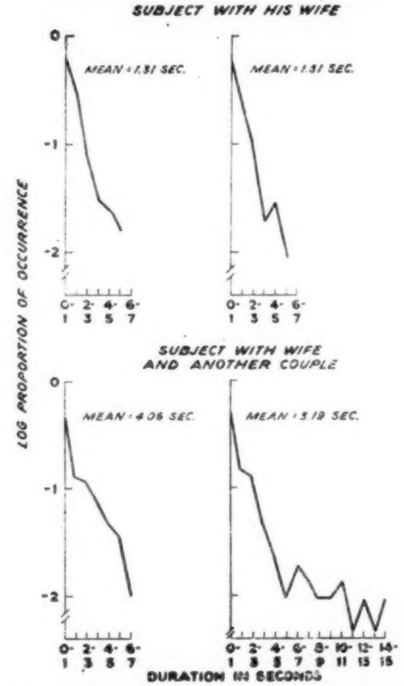


Fig. 3. Distributions of speech unit durations of one subject in four spontaneous situations.

this start effect may pass unnoticed, though it is always present and may itself prove to be a variable of some interest. As for the form of this particular distribution, beyond the start effect, the log plot at the bottom of Fig. 2 exhibits the scalloped discrepancy from linearity we would expect from a sample such as this, a miscellaneous combination of 51 different telephone conversations. The notch in the curve is also of interest and I shall comment on it below.

There are other situations where there are external constraints against short speech units, so that the start effect is greatly increased. These are situations in which the speech pattern is influenced by written material. We have previously published examples of data from situations where there existed such constraints (Starkweather, 1959; Hargreaves and Starkweather, 1959). Such data show an extreme reduction in shorter units compared to data from situations where there is no such external constraint.

But what of the ordinary conversation that we experience daily? Fig. 3 shows samples from four situations which occurred spontaneously in the course of one subject's day (Hargreaves, 1955). The subject and his wife had been continuously recorded for a period of two weeks, via a pair of small transmitters which they wore from morning to night. These four situations occurred on the fourteenth day of such recording. The two at the top of the figure were times when the subject was alone with his wife. In these samples, of about 150 speech units each, the discrepancy from a simple exponential distribution is not statistically significant. At the bottom of the figure are two situations during which the subject and his wife were sharing a table at mealtime with another couple with whom they were only recently acquainted. These samples, although about the same size as those above, clearly do not fit a single exponential. They have much larger mean durations because of an additional group of very long units. It is interesting to note that in these lower curves the sub-group of shorter units has in each case a slope essentially the same as in the two upper samples. It is as if in all four situations the subject emitted a group of units of one type, with a mean of about 1·3 seconds, while in the mealtime situations there was added another group of longer units.

Getting farther from such spontaneous situations, Fig. 4 shows data collected in role-playing situations (Starkweather, 1959). A standard role-player played opposite 20 different subjects under two different role instructions. Sample sizes were too small to treat data for individual subjects, so data were combined across subjects within each of the two situations. The standard role-player's units were similarly combined. Fig. 4 shows the composite distributions for the role-player and subjects in each situation. Again, a simple exponential provides a gross approximation to the data. There is, however, a slight scallop to each of the curves, one which is statistically significant with these large samples, so that we must reject the hypothesis that these samples were drawn from exactly exponential populations.

Investigators using the interaction chronograph method have frequently mentioned that the distributions of the durations of their units were characteristically J-shaped (Chapple, 1939 ; Goldman-Eisler, 1951 ; Saslow, Matarazzo and Guze, 1955). Chapple (1940) even asserted that the frequency distributions he obtained could be fit to the function:

$$F(t) = ae^{-bt} + ce^{-dt}$$

which is exactly the composite of two simple exponential distributions described above. Chapple published none of his data, so we have no indication of how adequate this formula was in actually describing his results. Guze and Mensh (1959), using a variant of the interaction chronograph method, also reported distributions of speech unit durations, but their sample sizes were unfortunately too small to foster much confidence in their conclusions about the form of the distributions obtained. Hess, Matarazzo, and Saslow (1960) have made a preliminary report of a study of speech unit durations. Group data were presented on the speech of 20 subjects in the non-stress periods of the standard interaction chronograph interview (Chapple, 1953). Results exhibit the typical scallop, but there is a more pronounced start effect.

The earlier studies of Starkweather and myself left in doubt whether the scallop effect results mainly from individual differences in expected duration, or from shifts in expected duration within each person's speech. Therefore, a systematic group of

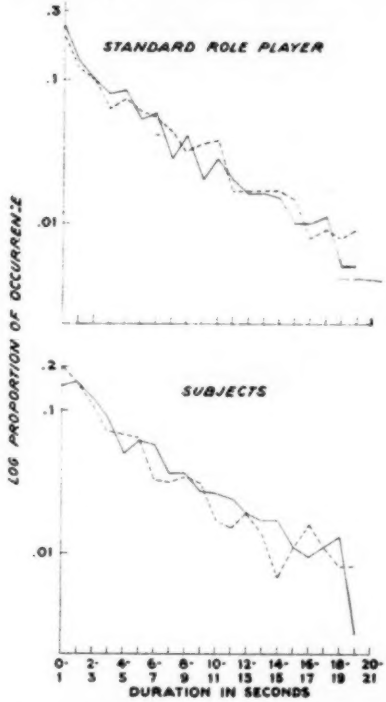


Fig. 4. Distributions of speech unit durations of standard role player and subjects under two role instructions.

speech samples was collected across both persons and situations, where samples of speech units were large enough to treat each person individually (Hargreaves, 1959). The subjects were ten college undergraduates, five pairs of dormitory room-mates. Speech samples were recorded in three situations: a free response period, a discussion period, and an interview. For the free response sample, a subject pair was recorded for an extended period during which they were alone together in their room, engaging in any activity they wished, but usually studying. This sample was a composite of many brief conversations with long periods of silence in between, occurring during eight or more hours of recording. The discussion situation was produced through Strodtbeck's technique of " revealed differences " (Strodtbeck, 1955). The room-mates each first filled out a multiple choice opinion questionnaire. They were then asked to resolve the questions on which their answers differed. In all five cases this produced a lively discussion, which was recorded for one hour. After these two samples had been obtained, each subject was interviewed individually for one hour by the experimenter. In the interview, the subject was asked to give an extensive, factual history of his life.

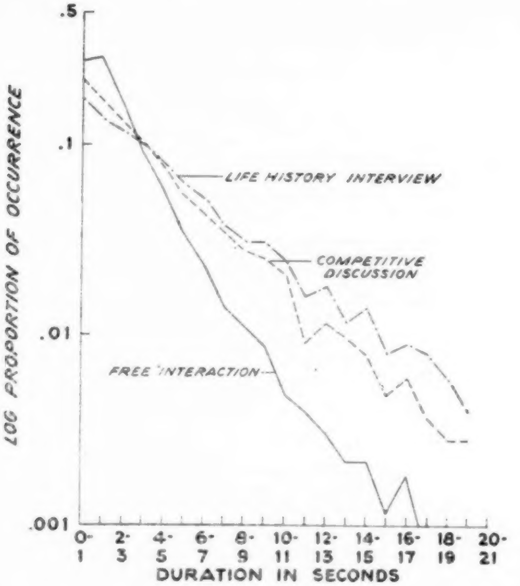


Fig. 5. Distributions of speech unit durations of ten subjects in three different situations.

While the subject was speaking, the interviewer made no verbal response and tried to restrict his non-verbal response as much as possible. Only when the subject could think of no more to say would the interviewer suggest another area he might discuss, and these brief conversational exchanges were later excluded from the speech sample.

As might be expected, the differences between these three situations produced large shifts in the mean speech unit durations for the individual samples. The free response utterances were shortest with a group mean of 2·58 sec. The discussion situation mean was 4·26 sec., while the interview utterances were longest, at 5·95 sec. These differences are highly reliable, the means for all 10 individual subjects showing the same progression across the three situations. Individual differences in mean duration were small relative to these situation differences.

To test whether the form of individual distributions varied, the raw curves were transformed to remove the effect of their differing means. With mean durations equated, heterogeneity χ^2 tests showed no evidence of other than random differences between the forms of individual distributions within each situation. Thus the combined curves from each situation shown in Fig. 5 accurately reflect the form of individual curves. Each of these group curves represents 3000 to 6000 speech units. All three curves appear to be roughly linear. The individual curves from the discussion situation were the closest, only three of the ten subjects showing significantly poor fits to a simple exponential. The free interaction data were the farthest from linear, showing a

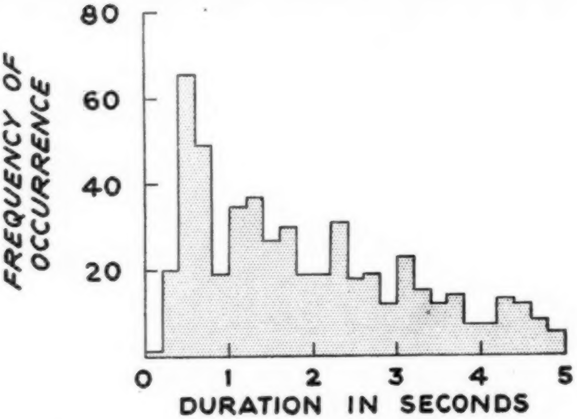


Fig. 6. Distribution of sound burst durations of one subject in an interview.

significantly poor fit for nine of the ten subjects. The discrepancy from linearity is of the two sorts we have discussed. There is a scallop to the curve, which presumably corresponds to some shift in the expected utterance duration within each individual sample. There is also an increased start effect in the free interaction situation. Remember that this discrepancy does not result from combining individual curves, but is a consistent effect throughout the samples from the ten individual subjects, and is specific to this situation in contrast to the other two.

This result led us to become interested in the start effect in its own right. Some suggestion as to what this increased start effect may represent is provided by the data in Fig. 6. While most of the studies reviewed above grouped data into one-second categories, this sample was measured to an accuracy of 50 msec. and presented here in 200 msec. categories. Furthermore, the units measured in this sample were the relatively continuous bursts of sound which make up speech. Separate units were counted when there was an intervening pause of about 50 msec. or longer. Previous studies have used unit definitions which group these basic sound bursts into longer units.[2] The sample in Fig. 6 is from the speech of a psychotherapist while having a leisurely discussion with his supervisor. In contrast to the data we saw before from telephone conversations, where the modal unit was only a quarter of a second long,

[2] *The Norwine & Murphy (1938) telephone study measured units consisting of all speech between successive responses of the listener. In the earliest study by the author (Hargreaves, 1955), units were similarly defined, except that a unit terminated if the speaker paused and there was no evidence he intended to continue within the next second or two. The interaction chronograph studies (Chapple, 1953; Matarazzo, et al., 1958) include gestural activity, and terminate a unit only when such activity ceases, as well as speech. In our more recent work (Starkweather, 1959; Hargreaves, 1959; Hargreaves & Starkweather, 1959) units were defined as ending when the speaker paused for one second or longer, without reference to content or gestural activity. Measurement of sound burst durations for the data in Fig. 6 were made from an oscillograph record of the rectified and filtered audio signal.*

here the modal unit is twice that, at a half second. But more interesting, there is a strong suggestion of a periodic effect in the distribution, as if his speech tended to come in multiples of some unit about 600 msec. long. Two sub-groups were selected from this sample and identified in the original recording. Twenty units were selected at the first mode, all of them 0·5 to 0·6 sec. long. All twenty of these were single monosyllables. Another twenty were selected at the second mode, 1·1 to 1·2 sec. long. Thirteen of these, or 65%, were pairs of monosyllables such as, " And, ah," or " Well, ah." The word rate within longer continuous bursts of speech is, for this man, considerably faster than two words per second. Thus these two modal sub-groups represent single or paired monosyllables which are elongated compared to the syllables, or even the words, in longer continuous speech units. A large proportion of such elongated and isolated monosyllables in a sample will reduce the total number of units under one second long. When units are tabulated into one-second categories, such a sample would show an increased start effect, but the periodicity would be obscured. This may account for the start effects seen in some of the samples reviewed above. The only other of the above studies where units were measured more finely than in one-second categories was the telephone study (Norwine and Murphy, 1938). The notch which is noticeable in Fig. 2 may have been another example of this sort of periodicity.

In summary, the studies reviewed in this paper suggest that a fairly accurate model for speech unit durations may be developed by considering speech as fitting a simple random decay process. A more exact model results from conceiving the expected utterance duration as shifting somewhat about the modal value reflected in the sample mean. While situational factors can have considerable influence on mean utterance duration, there seems also to be a relation between certain situational factors and the magnitude of the start effect. These effects of situation have been described for data now available, but the specific situational variables remain to be isolated.

References

Chapple, E. D. (1939). Quantitative analysis of the interaction of individuals. *Proc. Nat. Acad. Sci.*, **25**, 58.

Chapple, E. D. (1940). Personality differences as described by invariant properties of individuals in interaction. *Proc. Nat. Acad. Sci.*, **26**, 10.

Chapple, E. D. (1953). The standard experimental (stress) interview as used in Interaction Chronograph investigations. *Human Organiz.*, **12**, 23.

Goldman-Eisler, F. (1951). Measurement of time sequences in conversational behaviour. *Brit. J. Psychol.*, **42**, 355.

Guze, S. B. and Mensh, I. N. (1959). An analysis of some features of the interview with the Interaction Chronograph. *J. abnorm. soc. Psychol.*, **58**, 269.

Hargreaves, W. A. (1955). Time patterns in spontaneous conversation. M.A. thesis, University of Chicago.

Hargreaves, W. A. (1959). The duration of utterances. Ph.D. dissertation, University of Chicago.

Hargreaves, W. A. and Starkweather, J. A. (1959). Collection of temporal data with the Duration Tabulator. *J. exp. anal. Behav.*, **2**, 179.

Hess, H. F., Matarazzo, J. D. and Saslow, G. (1960). Characteristics of three interview interaction variables. Paper presented at Western Psychological Association, San José, California.

Matarazzo, J. D., Saslow, G., Matarazzo, R. G. and Phillips, J. S. (1958). Stability and modifiability of personality patterns manifested during a standardized interview. In Hoch, P. H. and Zubin, J. (Eds.) Psychopathology of Communication (New York).

Norwine, A. C. and Murphy, O. J. (1938). Characteristic time intervals in telephonic conversation. *Bell Sys. Tech. J.*, **17**, 281.

Saslow, G., Matarazzo, J. D. and Guze, S. B. (1955). The stability of Interaction Chronograph patterns in psychiatric interviews. *J. consult. Psychol.*, **19**, 417.

Starkweather, J. A. (1959). Vocal behaviour: the duration of speech units. *Language and Speech*, **2**, 146.

Strodtbeck, F. L. (1957). Husband-wife interaction over revealed differences. *Am. Soc. Rev.*, **16**, 468.

Verzeano, M. and Finesinger, J. E. (1949). An automatic analyzer for the study of speech in interaction and in free association. *Science*, **110**, 45.

MESSAGE UNCERTAINTY AND MESSAGE RECEPTION. II*

IRWIN POLLACK

Operational Applications Office, Air Force Command and Control Development Division, Bedford, Massachusetts

A previous experiment indicated (1) that the accuracy of message reception is relatively independent of the size of the message ensemble if the number of response alternatives is held constant; and (2) that the accuracy of monitoring performance is independent of the number of irrelevant response alternatives. The present study attempts to determine whether the same generalizations are warranted in the case where the message and response alternatives are randomly chosen from the English language. They are.

A previous study demonstrated that the accuracy of message reception was nearly independent of the size of the message ensemble, m, but was critically determined by the number of response alternatives available to the listener, r, where $r < m$ (Pollack, 1959). The previous study also questioned[1] whether this finding would be obtained had words been randomly selected from the entire language, rather than from a limited defined set (64 words in the first study). The present note tests this question.

The previous study also demonstrated[2] that monitoring performance was relatively independent of the number of alternatives other than the monitored alternative (range: 1 to 7 words). Is monitoring performance also relatively independent of the number of irrelevant alternatives if the irrelevant alternatives represent the entire remainder of the language ? The present note also tests this question.

* *This is Technical Note 60-23 of the Air Force Command and Control Development Division. This research supports Project 7682, Information Processing by the Human Operator, of the Air Research and Development Command Programme in Human Performance. The writer wishes to acknowledge the contribution of Mr. W. Sumby and Dr. C. Brown in discussing their research project which closely parallels the present study, and in sharing their stimulus materials with the writer.*

[1] *Pollack, 1959, Footnote 9, p. 1501.*

[2] *Pollack, 1959, Fig 4, p. 1502.*

Procedure

A pool of entries was obtained by tabulating the top entry from each page of a desk-size English dictionary. While entries were primarily single words, phrases such as *Rock of Ages* were also obtained.

Entries were selected from the pool either two-at-a-time with both entries recorded upon a single card ; or selected one-at-a-time with only a single entry recorded upon a card. Nearly 500 cards of each type were available for testing.

With cards of two entries, the talker read either the top or the bottom entry equally often, as determined by a scrambled programme. The listener's task was to respond whether the top or the bottom entry on the card had been read. These tests will be termed ' one-from-two alternative tests '.

With cards of only one entry, the talker read either that entry or some other entry equally often, again determined by a scrambled programme. The listener's task was to respond whether the recorded entry had been read (an acceptance response) or whether any other word had been read (a rejection response). These tests will be termed ' one-from-many alternative tests '.

The listener's criterion of acceptance or rejection is critical because he can arbitrarily choose to accept or reject a large percentage of trials. A four-interval confidence rating was employed in anticipation of the criterion problem. However, since a simple two-way split of the responses yielded nearly an equal number of acceptances and rejections, the confidence ratings will not be further considered.

Half of the listeners inspected the relevant card *before* the talker read the word in noise (pre-condition). Half of the listeners inspected the relevant card only *after* the talker read the word in noise (post-condition). The simultaneous listening under pre- and post-conditions provided an internal control that both conditions were simultaneously exposed to the same signal-to-noise (S/N) ratio.

A single talker was employed for all tests. The S/N ratios are based on VU readings upon the test words and true r.m.s. readings upon a white noise, band-limited between 100 and 6800 c.p.s. Listening was accomplished at a comfortable listening level. A trial block consisted of 24 items at a single S/N ratio. Successive items were read at 4 sec. intervals[3].

Two listening crews were employed: Crew 1 was a highly selected crew of three experienced listeners who had previously performed about 80 hours of listening in noise ; Crew 2 was a relatively inexperienced crew of six listeners.

Theory of the ideal observer

Since the ideal observer with perfect memory can faithfully reproduce the presented waveform, his performance should be equivalent for the pre- and post-conditions.

[3] *The manipulation of the cards within a short available interval probably worked to the slight detriment of the post-condition scores for the less experienced listening crew.*

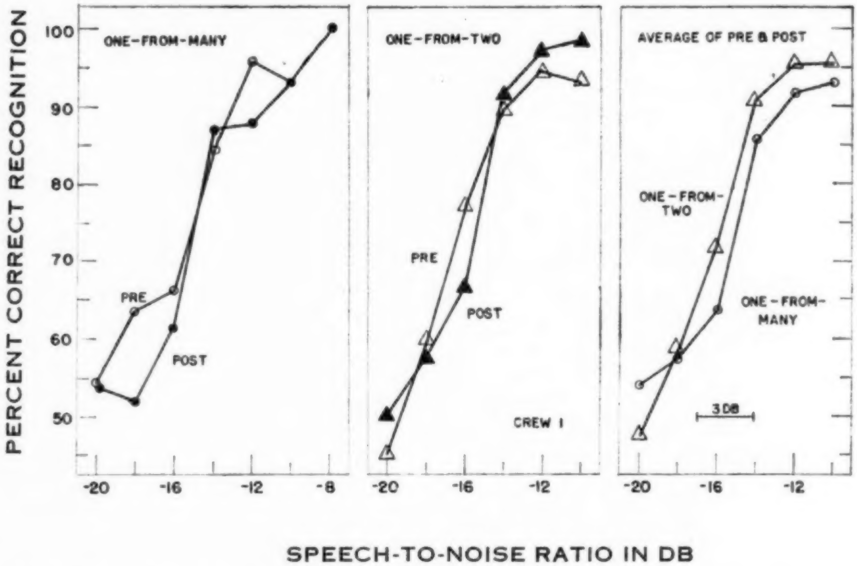


Fig. 1. Comparison of pre- and post-conditions in the one-from-many alternative tests and in the one-from-two alternative tests. Each point represents the results of two (at the extreme S/N ratios) to four (at the intermediate S/N ratios) 24-item tests with a crew of three experienced listeners.

However, for the ideal observer the performance in the one-from-two alternatives test should be substantially superior to the performance in the one-from-many tests. The theory of signal detectability predicts a 3 db difference in favour of one-from-two tests for the ideal observer on the assumption of equally-likely orthogonal messages (Tanner, 1960).

Results

The results for Crews 1 and 2 are presented in Figs. 1 and 2 in terms of the percentage of two-choice responses that were correct as a function of the speech-to-noise ratio. The first section compares the pre- and post-conditions in the one-from-many tests; the second section compares the pre- and post-conditions in the one-from-two tests; and the third section compares the two tests averaged over pre- and post-conditions.

With the exception of the one-alternative tests with the less experienced Crew 2, it is noted:

1. There is little systematic difference between the pre- and post-conditions, consistent over all tests.

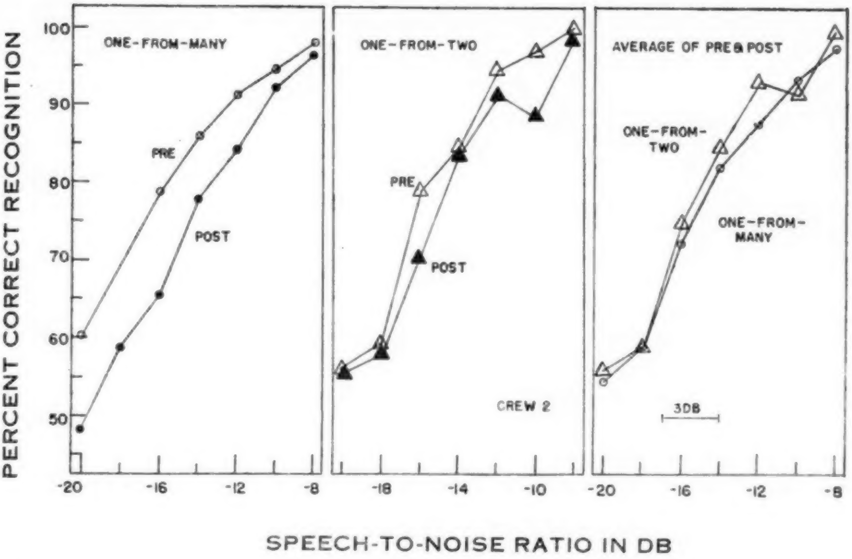


Fig. 2. Comparison of pre- and post-conditions in the one-from-many alternative tests and in the one-from-two alternative tests. Each point represents the results of two (at the extreme S/N ratios) to four (at the intermediate S/N ratios) 24-item tests with a crew of six relatively inexperienced listeners.

2. There is only a small difference in favour of the one-from-two tests over the one-from-many alternative tests. In any event, the observed difference is substantially smaller than the 3 db difference predicted by the theory of signal detectability for the ideal observer.

Conclusion

Knowledge of the class of possible messages before message presentation yields nearly the same intelligibility scores as knowledge of the class shortly after message presentation, when the time-of-occurrence of the message is known. This result, obtained with random selections from the dictionary, verifies an earlier finding obtained with a limited set of words. Contrary to the performance of the ideal observer, the performance of the human observer, under the conditions of the present study, does not suffer materially when called upon to monitor one from an extremely large number of alternatives rather than to select one from two alternatives.

References

Pollack, I. (1959). Message uncertainty and message reception. *J. acoust. Soc. Amer.*, 31, 1500.

Tanner, W. P., Jr. (1960). Personal communication.

PUBLICATIONS RECEIVED

Abstracts of English Studies, 3 (1960), 4-9.
American Annals of the Deaf, 105 (1960), 1-3.
Behavioral Science, 5 (1960), 2, 3.
ETC., 17 (1960), 2.
Journal of Speech and Hearing Research, 3 (1960), 2.
Leuvense Bijdragen, 49 (1960), 3/4.
Methodos, 11 (1959), 42.
Problems of Linguistics (Moscow), 9 (1960), 2-5.
Psychometrika, 24 (1959), 1-4 ; 25 (1960), 1.
Revista de Psihologie (Bucarest), 5 (1959), 1-4.
Volta Review, 62 (1960), 4-7.
Slama-Cazacu, Tatiana (1957). Relatiile Dintre Gindire si Limbaj in Ontogeneza (Academia Republicii Populare Romine, Bucarest).
Trojan, F. (ed.) (1960). Current Problems in Phoniatrics and Logopedics. Vol. I (S. Karger, Basel).
Wängler, H. H. (1960). Grundriss einer Phonetik des Deutschen (Elwert Verlag, Marburg).

SPEED OF UTTERANCE IN PHRASES OF DIFFERENT LENGTHS

I. Fónagy and K. Magdics
Hungarian Academy of Sciences, Budapest

This paper discusses the relation between the length of phrases and the speed of utterance. The dependence of speed on the length of phrases can well be expressed by means of exponential functions.

Previous investigations

Some prosodic features and even verse-rhythm itself have recently been derived from certain phonetic rules governing speech in everyday life (Vargyas, 1950, Szabédi, 1955). The "law of equalisation", that is the speaker's involuntary endeavour to pronounce short and long phrases in an approximately equal time (i.e., to pronounce long phrases more rapidly) plays an important part in these conceptions.

The relation between the length of words and sentences on the one hand, and speed on the other, was pointed out very early by the pioneers of phonetics. According to Sweet the diphthong /ei/ is shorter in *tailor* than in *tail*. According to Sievers, the speaker endeavours to pronounce phrases of somewhat different lengths each with an approximately equal duration.[1] Also Jespersen (1897) considers the speaker's endeavour to pronounce a longer sequence of sounds more rapidly than a shorter one, as a fundamental rule of prosody. Interposed clauses are pronounced particularly rapidly, at one go. Grégoire (1899) supported these untested observations by means of measurements. He gives the mean duration of /a/ in the French word *pâte* as 27 csec., in the word *pâté* as 20 csec.,[2] and in the word *pâtisserie* as 12 csec. According to Meyer (1904) quoted by Jespersen (1932), the mean value of the German /a/ sound in monosyllabic words is 13·2 csec., in disyllabic words 10·8 csec.; that of the long /a:/ in monosyllabic words is 26·5 csec., in disyllabic words, 22·0 csec. Some years later Meyer, co-operating with Gombocz (1909), investigated the quantity of the Hungarian vowels, and found the irregularities similar to those in German.

As these authors themselves emphasize, the available material is too small to draw far-reaching conclusions on the basis of the experiments. Further, we have to consider the fact that the words were not taken from spontaneous current speech.

After his examination of some Hungarian verse, Hegedüs (1934) came to the conclusion that the "quantity of sounds occurring between two 'centroids' (i.e., two

[1] *Vor allem aber regelt sich die Silbendauer zu einem grossen Teile nach der Silbenzahl der Sprechtakte denen die betreffenden Silben angehören. Sprechtakte, die an äusserem Umfang, d.h. eben an Silbenzahl, nicht zu verschieden sind, werden gern mit gleicher oder doch annähernd gleicher Dauer gesprochen, vgl. etwa Sprechtakte wie* heil, heilig, heiligere, *u.s.w.* (Grundzüge der Phonetik, *1885, Leipzig, 1901. 264 and k.*).

[2] *It cannot be left out of account that the /a/ sound of* pâte *was stressed, that of* pâté *generally unstressed.*

prominent syllables) is equalized by our rhythmical sense in a natural way." Hegedüs (1956) later studied the relation of phrase and speed on the basis of a larger amount of material. This time he took a more cautious stand on the problem. He merely stated that the shorter words and phrases were generally pronounced more slowly than the longer ones. At the same time he emphasized that this tendency was not always evident. Other writers, including, for example, Collinder (1939), have reached the same conclusion. Sovijärvi (1949, 1956) has recently dealt with the tendency to equalization in connection with sentence rhythm in Finnish and Swedish prose. Sovijärvi turns his attention to the syntactic units called " rhythmical periods " (rytmijakso), instead of the phrase. The " rhythmical period " is an independent unit of thought (Gedankenschritt), a " mental breathing " containing several phrases (puhetahti). According to the author only the " rhythmical period " can be considered as a primary rhythmical unit. The common feature of the phrase (puhetahti) and the " rhythmical period " (rytmijakso) is the tendency towards equalization of quantitative differences.

Experiments

Considering the importance of the observed " law " and the far-reaching conclusions drawn from it, experiments on the theory of equalization are still rather rare.[3] One of us has measured the length of the phrases of some poems, and determined the speed as a function of the length of phrases (Fónagy, 1959). By *speed of utterance* we mean the number of sounds uttered in a second. By *phrase* we mean—here and in the following pages—a rhythmical unit, a sound sequence governed by a stressed syllable (generally the first syllable of the phrase). The unity of the phrase is underlined by the rising pitch in the stressed syllable and the falling intonation of the rest of the phrase. The phrase cannot be interrupted by any pause. If the speaker breaks off the sentence at an unusual place, within a section that would naturally be regarded as a phrase-unit, we consider the section as two different phrases. In some exceptional cases the phrase may be shorter than a word. One subject, in making a recording, said " *Akármi*lyen *ros*szul is *hang*zik " /'aka:r'mijan 'rossul iʃ 'hangzik/ (however badly it may sound), giving a great emphasis to the third syllable of the first word. The sentence is clearly divided into four phrases, like a musical rhythm of four beats.

The results of the investigation supported the hypothesis only to a certain extent. The speakers uttered phrases consisting of 5 to 6 sounds really more slowly than the longer ones. The " rule of equalization " however rarely affected phrases of more than 6 syllables (*op. cit.*, 182).

The material examined consisted exclusively of poems, so we must not take it for granted that the above regularities prevail in other situations, for instance, in the course of a conversation, as well. We decided, therefore, to examine the relation of the length

[3] *In his article on the speed of utterance Hegedüs (1953) determined the speed word by word; in his paper on the phrase, he determined the duration of the sections of speech limited by pauses. Therefore it would be difficult to employ his material for investigation of the speed of phrases.*

of phrase and its speed on the basis of more varied material. We supplemented the poems with tales (1,656 syllables), scientific and literary lectures (2,297 syllables), sports transmissions (1,435 syllables), recordings of spontaneous conversations (2,375 syllables) and of the speech of children at the age of 5 - 6 years (2,312 syllables).

The results of the investigation are summarized in Figs. 1 - 7. A considerable decrease of speed is to be found only in the shortest syllables. The duration of phrases of different lengths is by no means the same. Long phrases often have double the duration of short ones, both in spontaneous conversation and in reading.

There are certain differences between the different types of material. The curves for poems (Fig. 1) and tales (Fig. 3) fall most steeply. This means that in certain situations, where the musical or aesthetic view-point prevails, the " tendency to equalization " makes its effect vigorously felt in the short phrases. In the tales, the average duration of a sound rose from 8 csec. to 16 csec. according to the length of the phrase. The extreme values were closer to each other in conversation where in long phrases the average length of a sound was 6 csec., in the shortest ones, 11 csec. The curves for the tales (Fig. 3) and for reading prose (Fig. 2) fall steeply showing that in these cases the speed of phrases consisting of more than two or three sounds suddenly rises. In spontaneous conversation (Fig. 5) the acceleration is more gradual.

In one of the tales analysed, some verse-lines were hidden. The phrases rhyming with one another—and consisting of approximately the same number of syllables—sometimes had the same, or nearly the same, duration.[4]

Van itt ész	hamar kész	uram király,	ide nézz !
102 csec.	102	107	95
Vászonkötós kukta	lyukas vödröt hozott,		
140 csec.	145		
Könnyeit a király	abba eregette.		
123 csec.	110		
Már az egész várost	árviz fenyegette.		
137 csec.	137		

The " tendency to equalization " is operative in the traditional turns of story-telling too.

. . . . Nekeresden	túl volt	Nevesincsen	innen volt
90 csec.	95	100	95

Perhaps these correspondences are not altogether accidental, although their occurrence is very rare.

The average speed of utterance (being independent of the length of phrases) similarly differs according to the various types of material (see Table 1). Reciting poems was found to have the lowest, sports transmission the highest speed. It is interesting that the speed of the sports transmissions is hardly greater than that of the conversations. The result would probably change if we based our determinations of the average speed on a whole conversation or an entire transmission considering the pauses as well.

[4] *Hegedüs (1956) found that in poems the rhyming phrases at the end of the lines rhyme with one another from the view-point of duration as well.*

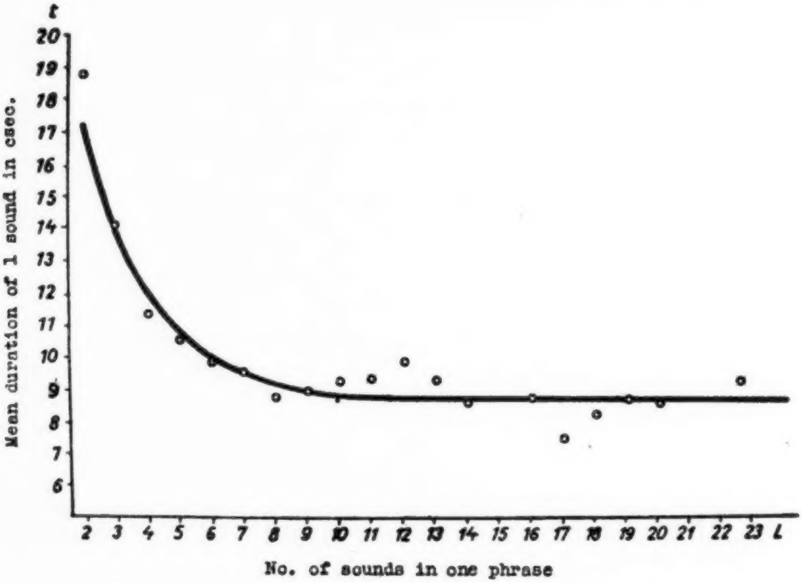


Fig. 1. The relations between length of phrase and mean duration of the sounds in poems. L = the number of sounds in one phrase; t = mean duration of one sound in centiseconds. The small circles denote the results of measurements. The curves of Figs. 1 - 7 can be represented by the equation $y = a + be^{-cx}$ (see footnote 5).

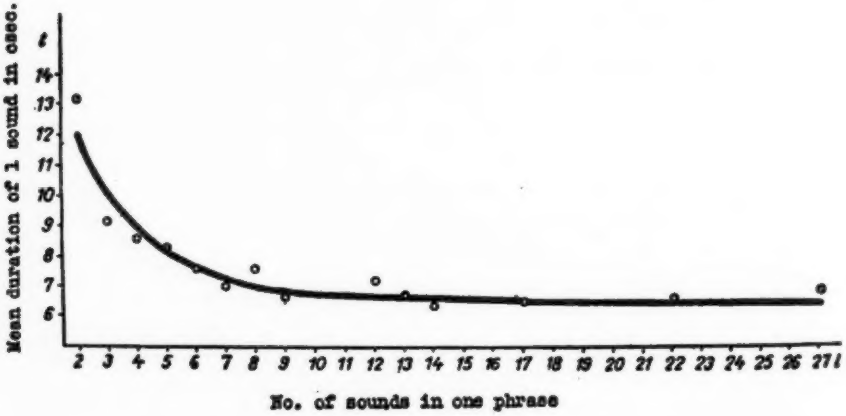


Fig. 2. The relation between length of phrase and mean duration of the sounds in prose reading.

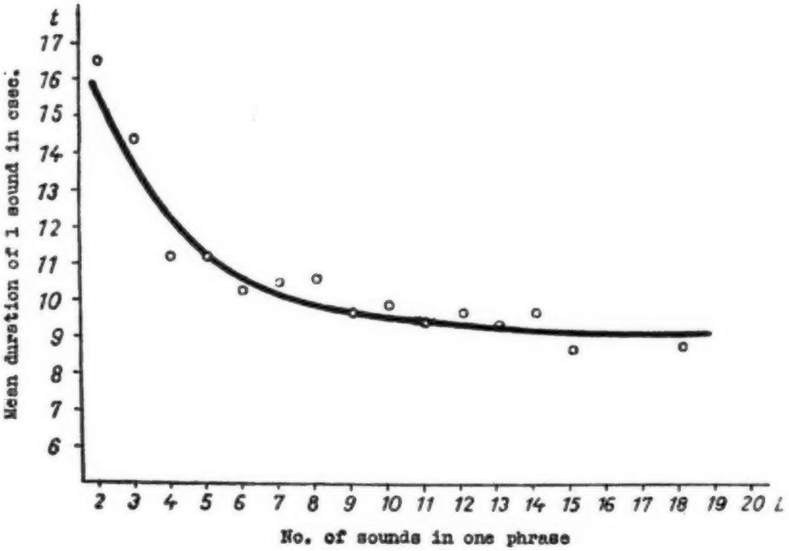


Fig. 3. The relation between length of phrase and mean duration of the sounds in tales.

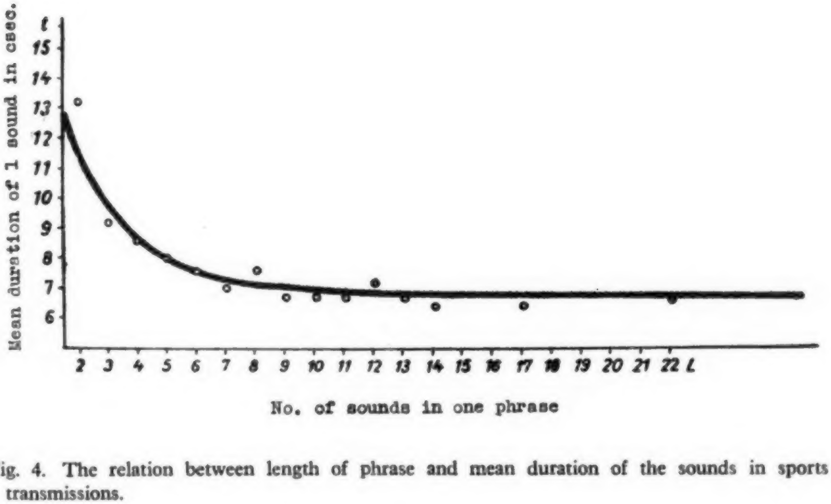


Fig. 4. The relation between length of phrase and mean duration of the sounds in sports transmissions.

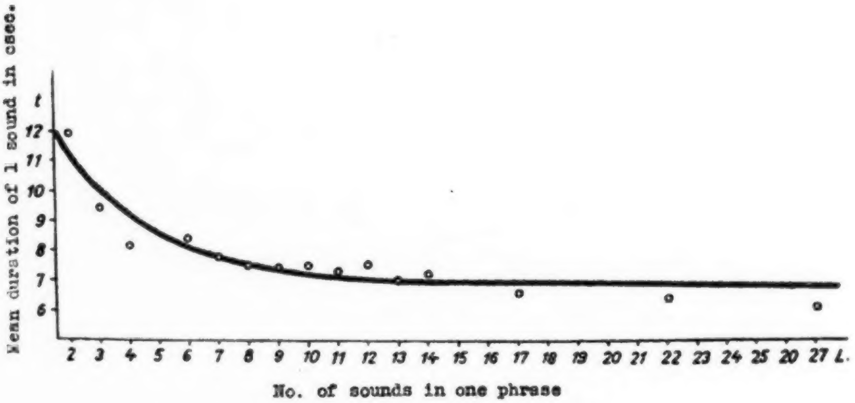


Fig. 5. The relation between length of phrase and mean duration of the sounds in conversation.

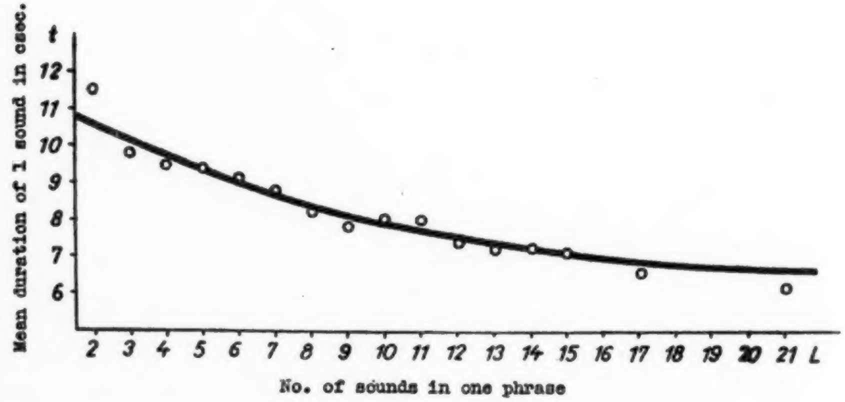


Fig. 6. The relation between length of phrase and mean duration of the sounds in children's speech.

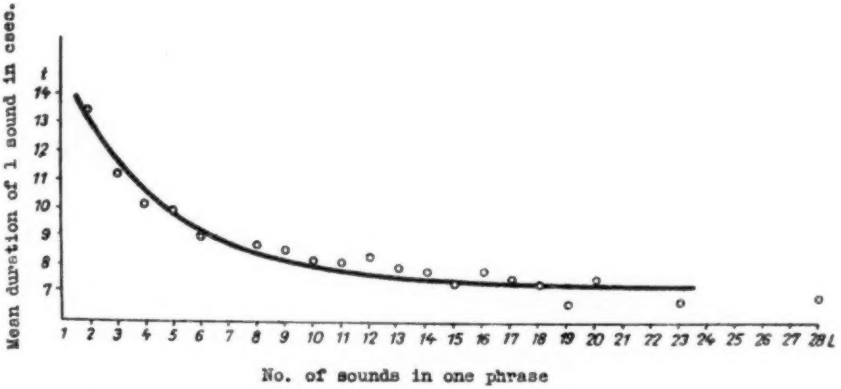


Fig. 7. This figure gives a summary of the relation between length of phrase and mean duration of sounds for all the material examined.

The sports commentator is able to keep pace with the events in his reporting because he omits the pauses, and not—or not in the first place—because he shortens the duration of the sounds. The pauses are much more " elastic ". Their length is reduceable to a greater extent, whereas the speech sounds cannot be " compressed " beyond a certain limit. The mean value of the speed of utterance in Hungarian, according to our investigations, is 11·35 sounds per second.

The exponential law

Though the differences of duration are not equalized phrase by phrase, speed depends to a certain extent on the length of the phrase. The relation between speed and duration may be expressed by an exponential equation. Our data are closely approached by curves of the equation:

$$y = a + b\, e^{-cx}$$

(cf. Figs. 1 - 7)[5]

It is not surprising that the inherence between speed and length of phrase can be approached by means of an exponential curve, if we consider similar relations observed

[5] *The curves of the following equations occur in the given figures :*

Fig. 1 (poems): $y = 21{\cdot}6e^{-0{\cdot}47x} + 8{\cdot}7$
Fig. 2 (prose reading): $y = 12{\cdot}24e^{-0{\cdot}40x} + 6{\cdot}5$
Fig. 3 tales: $y = 12{\cdot}88e^{-0{\cdot}36x} + 9{\cdot}1$
Fig. 4 (sports transmissions): $y = 11{\cdot}94e^{-0{\cdot}48x} + 6{\cdot}8$
Fig. 5 (conversation): $y = 8{\cdot}41e^{-0{\cdot}32x} + 6{\cdot}8$
Fig. 6 (children's speech): $y = 5{\cdot}64e^{-0{\cdot}12x} + 6{\cdot}1$
Fig. 7 (summary): $y = 9{\cdot}76e^{-0{\cdot}30x} + 7{\cdot}2$

TABLE 1

	Mean values of the speed of utterance (sounds per second)
Poems	9·40
Tales	9·57
Prose reading	10·73
Children's speech	11·68
Conversation	12·89
Sports transmissions	13·83
Average speed	11·35

Mean values of the speed of utterance in different types of material.

in the different fields of physiology. According to Janisch (1927) the exponential law is of basic importance in the biosphere, and may be operative in other areas of the world of nature as well. Weber-Fechner's law determining the logarithmic character of our sensation is, according to the author himself, only a special case of the exponential law.

STRESS, BREATHING AND SPEED OF UTTERANCE

The relationship between speed and phrase-length is by no means explained by connecting it with analogous biological processes. It is obvious that speech-sounds cannot shorten beyond a certain (undetermined) limit, and " saturation " follows necessarily. The sound can neither be pronounced nor perceived beyond a certain limen. The more we approach this (unknown) value, the less the acceleration. Nevertheless it needs to be explained why only the sounds of phrases consisting of two or three syllables lengthen to a considerable extent.

The slowing down of the short phrases becomes perhaps more easy to understand by reversing the causal relation. Not only does the speed depend on the length of the phrase, but also the length of the phrase may be influenced by speed. The story-teller sometimes utters a word with particular emphasis, slowly and carefully articulating the sentence:

Száz	palotámban	ezer	terem. . . .
(In hundred	castles of mine	[are] thousand	rooms)
45 csec.	82	85	75

The average duration of the short phrases is often lengthened by interjections like *hát, hej* (well, hey !)

According to our definition, phrases extend from one stressed syllable to another. The stressed syllables are, as is well-known (Panconcelli-Calzia, 1917, Fry, 1955) longer than unstressed ones. In a phrase of two syllables, half the phrase is lengthened by stress, in a phrase of three syllables, a third part, etc. The longer the phrase, the smaller the lengthening effect of stress. Phrases of 1 to 5 syllables were more usual in the tales than in conversations, or in sports transmissions. Despite this fact—and even

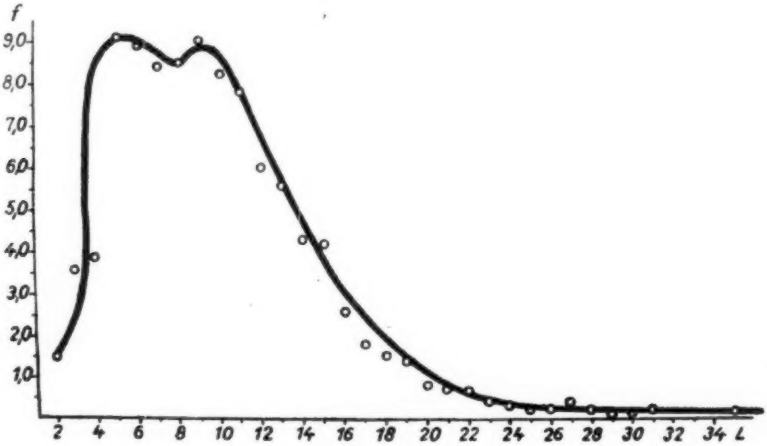


Fig. 8. Frequency distribution of phrases of different length. L = the number of sounds in one phrase ; f = the frequency of phrases of different length expressed as a percentage. The two peaks of the curve are significant, for they occur in the same place in every type of material. This is probably due to the structure of Hungarian syllables. Disyllabic words consisting of 7 sounds are less frequent in Hungarian than those consisting of 5 to 6 sounds, and trisyllabic words consisting of 8 to 9 sounds are more frequent than those consisting of 7 sounds.

partly because of it—the average duration of phrases in the tales was greater than in conversations or sports transmissions (cf. Figs. 3, 4 and 5). This may also help to explain the fact that the curve of the tales is more abrupt than that of the sports transmissions.

Nevertheless the lengthening caused by stress is not considerable ; the phrases *hej, hát* are comparatively rare, so that it is hardly possible to explain in this way why the accelerating tendency relaxes in phrases longer than four syllables.

Perhaps we can approach the solution of the problem better by setting out from the physiology of breathing. Inspiration and expiration are rhythmical processes, and it is possible that the speed of phrases somehow adjusts itself to this physiological rhythm.

Our records hardly give any information about the technique of the speakers' or readers' breathing. We have therefore supplemented our investigations by asking ten speakers (all having a university degree or matriculation) to read aloud part of a tale. Parallel with the recording we registered the process of breathing by means of a pneumograph. We were now able to establish the sections delimited by inspiration as well as the phrases determined by the incidence of stress.

Every inspiration coincided with some natural phrase-limit in the case of reading alcud. Several phrases, or sometimes even several sentences, were read out in one breath. Inspiration took place in 87·5% of the cases after the end of the sentence ; 9·4% of the breath-divisions coincided with the beginning or end of a clause, 3·1%

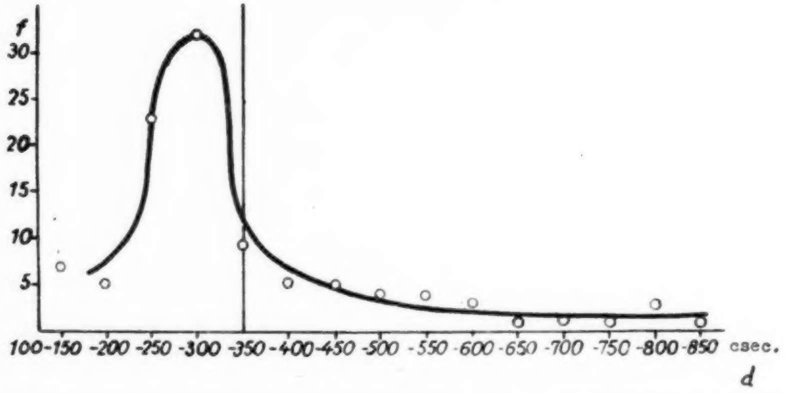


Fig. 9. Frequency distribution of the durations of expiration during speech. d = duration of expiration in csec. grouped into the classes 100 - 150 csec., 151 - 200 csec., etc. f = frequency expressed as a percentage of durations falling within each class. The vertical line indicates the mean duration of mute expiration (which generally fluctuates between 300 and 350 csec.).

with some other phrase-limit. The duration of an expiration during speech was variable, the individual averages fluctuated between 233 csec. and 490 csec. One and the same speaker took breath sometimes after 150 csec., sometimes after 716 csec. This strong fluctuation is diametrically opposed to the great regularity of silent breathing. The individual average of (mute) respiration fluctuated between 312 csec. and 370 csec. in the course of the experiments. In spite of this, the mean value of the duration of expiration during speech approached the mean value of mute expiration. Between two inspirations, 339 csec. passed on an average in the course of the mute-breathing experiments, 321 csec. in the course of reading aloud. This means that despite the large range of variation (cf. Verzeano, 1950) according to the complexity of concepts and other variables of communication, the physiological needs—the tendency to least effort—are still dominant. The slight shortening of the breathing periods in speech is due probably to increased mental and muscular activity (see Fig. 9). With a knowledge of the average duration of (mute) expiration, the mean value of expiratory sections in speech can be predicted.

But it is not yet proved that the regularities of expiration during speech can be traced back in every case to relatively simple physiological factors. It has to be tested, for instance, whether the mean value of the length of phrases can be determined on the basis of the respiratory mechanism. The majority of phrases consist of 6 to 10 sounds, that is, approximately of 4 to 7 syllables, the mean value of their duration is 79.3 csec. A certain fluctuation is found according to the different type of spoken material. In conversations the phrases are considerably shorter than in prose readings (see Table 2). If we compare the mean value of phrase-duration with that of the expiration period (cf. Table 2 and Fig. 9) we see that one breath includes about 4 phrases. There is no evidence that the physiology of breathing could explain why

TABLE 2

Number of sounds	Poems	Tales	Prose reading	Children's speech	Sports transmissions	Conversation	Mean value
	%	%	%	%	%	%	%
1 - 5	15·75	23·03	8·80	24·91	16·85	20·82	18·36
6 - 10	50·13	49·72	40·77	40·00	47·19	43·68	45·25
11 - 15	28·08	21·35	32·00	28·64	24·72	25·09	26·65
16 - 20	5·00	5·62	10·30	4·10	8·71	7·34	6·84
21 - 25	1·00	0·28	4·30	2·10	1·12	2·56	1·89
26 - 30	—	—	2·57	0·17	1·40	0·51	1·16
31 - 35	—	—	1·28	—	—	—	1·28
Mean value of duration of phrases in csec.	87·32	84·41	92·55	65·34	68·20	72·10	78·32

The length and duration of phrases in different types of material.

the breath-period is divided into four phrases instead of two, or why it is divided into phrases at all.

It might be supposed that the longer phrases are uttered more rapidly because the speaker is afraid of running out of breath. But in this case it is not clear why it is just in the longer phrases that the relation between the number of syllables and speed is blurred. We can demonstrate the incorrectness of our supposition in a simpler way. If our supposition were correct, the speaker would utter the words in the longer expiratory sections more rapidly. In reality the speed is quite independent of the length of the breath-period. The mean duration of a sound in the different sections fluctuates between very narrow limits (9·1 csec. - 9·7 csec.) and even this slight fluctuation cannot be correlated with the length of the breath-period (see Fig. 10). Moreover, if the supposition were correct, the phrases directly preceding the inspiration would have to be uttered more rapidly than the phrases directly following it. But this is in fact not so. The mean duration of phrases preceding inspiration is 9·45 csec., in phrases following it 9·35 csec. This difference of 1 msec. can hardly be regarded as significant.

SPEED OF UTTERANCE AND THE REPETITION COMPULSION

The tendency towards the equalization of phrases cannot be derived from the physiology of breathing. The explanation must be found on a higher level.

We have to reckon, we suspect, with a psychological factor: the so-called sense of rhythm, with the repetition-compulsion whose purpose is to free the organism from stimulation and restore it to a state of equilibrium (Freud, 1940). Periodicity of time-pattern, and other periodicities in speech, are domesticated forms of this compulsion. They are combining new information carried by new words with the repetition of certain rhythmic structures. (The domesticated form of the repetition-compulsion has been designated as the " order-principle " by Hermann, 1922.)

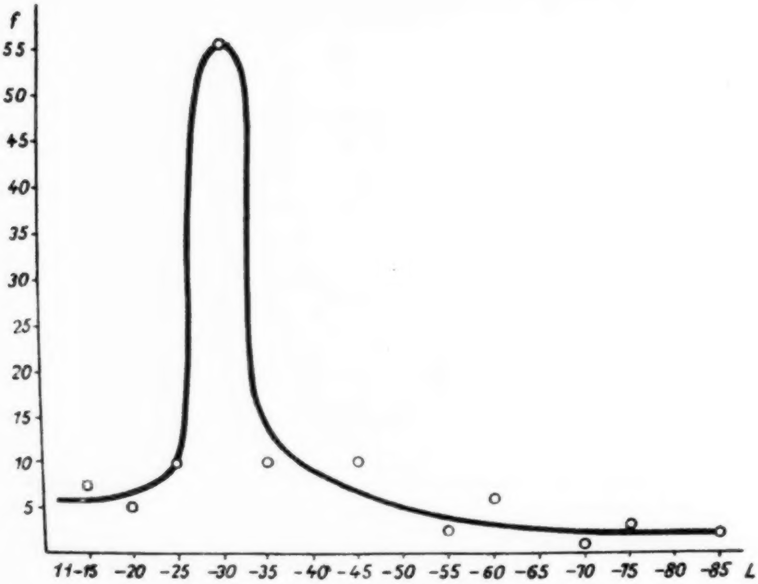


Fig. 10. Frequency distribution of the number of sounds uttered in one expiration. L = number of sounds grouped into the classes 11 - 15 sounds, 16 - 20 sounds, etc. f = frequency expressed as a percentage of expirations falling within each class.

The tendency towards the equalization of phrases of different lengths should be considered as a manifestation of the repetition-compulsion (or " order-principle "). Shorter phrases seem to be more subjected to this compulsion since a change of duration is more easily perceived in a phrase of 2 to 6 syllables than in a 10 or 20-syllable one. In the former case also the objective proportion changes to a greater extent. Moreover differences can be more easily kept in evidence on a lower level. We know that our senses react to stimuli on the whole according to a logarithmic scale. It is possible that we appreciate also the changes of duration in such a way, and are more sensitive to changes between 30 csec. and 60 csec. than to those between 60 csec. and 120 csec. It is also obvious that in the course of reciting or story-telling the tendency towards equalization and regularity operates more strongly, as the repetition-compulsion is more prevalent in poetry and literature.

Thus we have returned to the starting point of our investigations. The aestheticians are right in connecting verse-rhythm and the operation of the " equalization law " in speech. This tendency can be clearly proved under certain circumstances (in the case of short phrases). Nonetheless, the hope of tracing back poetic rhythm and the sense of rhythm to linguistic features or to the physiology of breathing seems unjustified. On the contrary: the tendency to equalization mentioned is explicable only by reference to " the sense of rhythm ", the " order-principle " as already suggested.

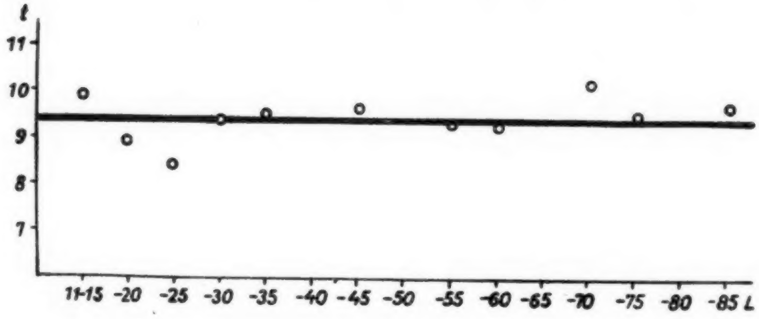


Fig. 11. Rate of utterance as a function of the number of sounds uttered in one expiration. L = number of sounds grouped into the classes 11 - 15 sounds, 16 - 20 sounds, etc.; t = mean duration of one sound. The curve does not show any tendency; the slight fluctuation is due to the small number of samples within the various categories. The value of t for the most "charged" group consisting of 26 - 30 sounds (which represents 55% of the cases, cf. Fig. 10) falls on the horizontal line indicating the mean of the mean durations for different types of material (cf. Fig. 7).

Conclusions

Our observations can be summarized as follows:

1. The speed of utterance differs according to the various types of material. In reading poems, 9·4 syllables are uttered in a second, in sports transmissions 13·83 syllables. The average value of the speed of utterance, according to our records, is 11·35 sounds per second (see Table 1).

2. Stress and intonation reduce the sentence to phrases, to units generally larger than the word. The length of phrases varies a great deal. According to the material examined, the shortest phrase consisted of 2, the longest of 35 sounds. Phrases of 6 and 9 sounds (of 2 or 3 syllables) are the most frequent (see Fig. 8). The duration of the phrases fluctuated between 45 csec. and 272 csec., the mean value of their durations was 78·3 csec. (see Table 2).

3. The speech-flow is divided into sections also by inspiration. The lengths and durations of expiratory sections vary during speech. The duration of the shortest section was 117 csec., the number of sounds, 14. The duration of the longest section was 854 csec., the number of sounds was 84.

4. The distribution of expiratory sections is even more concentric than that of phrases (cf. Figs. 8 and 10). The mean value of the expiratory sections in speech is somewhat smaller than that of the mute expiratory sections (see Fig. 9).

5. As many as 87·5% of the inspiratory pauses took place at the end of the sentences, 9·4% after clauses, 3·1% fell at other phrase-limits.

6. The speed of expiratory sections is entirely independent of the length of the sections (see Fig. 11).

7. The speed of the phrases depends on their length. The shorter phrases are uttered more slowly. The dependence of speed on length gets much less, however, in phrases consisting of more than 2 or 3 syllables. The lengthening is considerable only in quite short phrases. The dependence of speed on the length of phrases can be well expressed by means of exponential functions (see Figs. 1 - 7).

8. The length and average duration of phrases, the speed, the interconnection of length of phrases and the speed is more or less variable with different kinds of spoken material (poetry, prose, conversation, etc.; see Figs. 1 - 6).

9. The increase of speed in the longer phrases is independent of the regularities of breathing. The tendency to equalization is a manifestation of the repetition-compulsion.

References

Collinder, B. (1939). Das Wort als phonetische Einheit (Uppsala), 66.

Fónagy, I. (1959). A költói nyelv hangtanából (Budapest), 180.

Freud, S. (1940). Jenseits des Lustprinzips. Gesammelte Werke, XII (London), 21, 38.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *J. acoust. Soc. Amer.*, **27**, 765.

Grégoire, A. (1899). Variation de la durée de la syllabe français. *La Parole*, **1**, 161.

Hegedüs, L. (1934). A magyar nemzeti versritmus kérdése (Pécs), 15.

Hegedüs, L. (1956). Sprechtempoanalysen im Ungarischen. *Zeitschr. f. Phon.*, **10**, 13.

Hermann, I. (1922). Randbemerkungen zum Wiederholungszwang. *Int. Zeitschr. f. Psychoanalyse*, **8**, 12.

Janisch, E. (1927). Das Exponentialgesetz als Grundlage einer vergleichenden Biologie (Berlin).

Jespersen, O. (1897-9). Fonetik (Copenhagen), 508.

Jespersen, O. (1932). Lehrbuch der Phonetik (Leipzig, Berlin), 180.

Meyer, E. A. (1904). Zur Vokaldauer im Deutschen. *Nordiska Studier tillegnade A. Noreen.*

Meyer, E. A. and Gombocz, Z. (1907-8). Zur Phonetik der ungarischen Sprache. *Le Monde Oriental*, **2**, 20.

Panconcelli-Calzia, G. (1917). Über das Verhalten von Dauer und Höhe im Akzent. *Vox*, **27**, 127.

Sievers, E. (1901). Grundzüge der Phonetik (Leipzig), 264.

Sovijärvi, A. (1949). Huomioita puherytmiikasta. *Virittäjä*, **53**, 122.

Sovijärvi, A. (1956). Über die phonetischen Hauptzüge der finnischen und ungarischen Hochsprache (Wiesbaden).

Szabédi, L. (1955). A magyar ritmus formái (Bucharest), 20.

Vargyas, L. (1955). A magyar vers ritmusa (Budapest).

Verzeano, M. (1950). Time patterns in speech for normal subjects. *J. Speech & Hearing Disorders*, **15**, 199.

THE ORGANIZATION OF A RUSSIAN-ENGLISH STEM DICTIONARY ON MAGNETIC TAPE

D. W. Davies

National Physical Laboratory, Teddington

The organisation of a dictionary on magnetic tape is described. A single pass of the dictionary tape is sufficient to process a block of text words without reverse motion. Most Russian words are represented in all their inflected forms by a single entry in the dictionary. Even irregular nouns can usually be represented by two entries, but many verbs require two or three entries. The space occupied by an entry is very little greater than that required for a single inflected form. Mobile vowels are treated with very little extra space or complication of programme. The form of the entries is designed so that a simple technique for interpreting the inflections can be used. This is outlined at the end of the paper.

In a computer programme for translating from Russian to English, one of the first steps is to look up each word of the Russian text in a Russian-English Dictionary. This must be a special dictionary, accessible to the computer. It is important to make the dictionary as comprehensive as possible, because missing words leave gaps in the syntactic information, and may make the analysis of the Russian sentence ambiguous. An adequate dictionary for a particular branch of science might have 100,000 entries. A dictionary entry consists of two distinct parts. One part is the Russian word, which is the 'key' to the dictionary, enabling the computer to identify the entry it needs. The other part is the output of the dictionary, the English correspondents of the Russian word, its use in idioms, special functions in a sentence, government of other words, and so forth. In this paper we shall mainly be concerned with the problem of identifying the dictionary entry. An estimate of the size of an entry is perhaps necessary. There are good reasons for employing variable sizes of entries, but we will estimate that on average they occupy 1,500 bits of storage space.

In the dictionary organization which we shall describe, the identifying data or 'key' is more than just a Russian word, it is a 'stem', and 'paradigm'. This accounts for the non-triviality of the identification problem. The idea of a stem and paradigm will be explained in the next section of the paper.

The estimated storage space required by the whole dictionary is 150 million binary digits. These amounts of data are usually stored on magnetic tape, because it is the only readily available store which is cheap enough to be considered. Magnetic tape has the property that it can give up its data at a high rate, perhaps half a million binary digits per second, but only in the sequence in which it has been stored on the tape. All the data in the dictionary could in principle be stored on one reel of tape. The reading of this, from end to end of the tape, would perhaps take 5 minutes. It is obviously important to look up as many words as possible during this reading of the dictionary tape. The method of doing this is well-known, but it will be described briefly, for completeness.

The entries on the tape are assumed to be in alphabetic sequence with regard to their Russian words. The words in the text to be translated are first rearranged in alphabetical sequence. A magnetic tape containing these words in alphabetic sequence can be read on one tape transport, while the dictionary is being read on another transport. Programming of the start and stop of the transports is arranged so that the words being looked for correspond to that part of the dictionary which is currently being examined. (If words being looked for occur at a low density in the dictionary, each one may be held in store for a time while the computer is 'thumbing through' the dictionary to the appropriate place.) The dictionary outputs for these words are put out by the computer on a third tape transport. These must now be rearranged back to the order of their key words in the original text. For this purpose we will have assigned serial numbers to the words in the original text, and transferred these numbers to the corresponding dictionary output so that at the last stage, rearrangement according to these serial numbers reconstitutes the text-sequences.

By chosing a suitable amount of text for processing at one time, this procedure can be quite rapid, but its use results in a limitation on the mechanical translation procedure as a whole. In some mechanical translation procedures that have been described, a word may be looked up several times in the dictionary, with variations such as the removal of prefixes. This would not be economic with the 'sort and merge' scheme of look-up. If multiple look-up cannot be avoided, it should be done on a limited scale, perhaps combining the second stage of look-up for one block with the first stage for a new block of text. In general, it is better to get from the dictionary at one time all the information that could possibly be wanted, even if some is of doubtful relevance.

Special purpose computer stores can be made (because of the relatively unchanging nature of the dictionary data), which give random access to all the items in a small fraction of a second. With these stores, the problems dealt with in this paper are made much easier, but the problem of keeping the dictionary up-to-date is made more difficult. We shall be concerned only with the well-tried and generally available store, which is magnetic tape.

The need for a dictionary of Russian stems

A Russian word may be capable of taking on a number of different forms by changing its last few letters. Russian nouns, for example, can take any of six cases and either the singular or plural number. These 12 forms are not all different, but there are about 10 different inflected forms in general. The word СТОЛ can appear as:

	Singular number		Plural number	
Nominative case	СТОЛ	(table)	СТОЛЫ	(tables)
Accusative case	СТОЛ		СТОЛЫ	
Genitive case	СТОЛА		СТОЛОВ	
Dative case	СТОЛУ		СТОЛАМ	
Locative case	СТОЛЕ		СТОЛАХ	
Instrumental case	СТОЛОМ		СТОЛАМИ	

A table like this is called a 'paradigm' because it forms an example for the declension of a large class of nouns similar to СТОЛ. For our purposes, however, the important part of the table is the set of endings which are attached to the stem СТОЛ-. We shall use the word 'paradigm' to denote a table of endings, in which the ending for each case and number of the word is described. The ending for the nominative and accusative singular in the case of СТОЛ is null. We shall represent this conventionally as *.

Adjectives have not only case and number, like nouns, but also variable gender, and a short form comprising four inflections. Verbs have a completely different inflectional scheme based on mood, voice, tense, person, and number, but not all these in all combinations. To avoid circumlocution we may refer to case and number, using nouns as the concrete example, where in general the idea to be conveyed is that of 'a slot in the paradigm table'.

The dictionary could contain a separate entry for each of the inflected forms, but this is plainly redundant, because most of the information in these entries would be the same. The only information which varies is the identification of case and number. A dictionary containing a separate entry for each inflected form is called 'fully inflected'. A great saving in storage space can be achieved by making a single entry serve for all the inflected forms of a single word. It must contain the stem, or unchanged paıt of the word (СТОЛ-) and an indication of the set cf permissible endings (*, -А, -У, -Е, -ОМ, - Ы, -ОВ, -АМ, -АХ, and -АМИ) with their respective cases and numbers. A dictionary containing this type of entry is usually called a 'stem dictionary', but in it we have not only a stem, but also a set of endings, or some coding which implies these. A better name would therefore be a 'stem-paradigm dictionary'.

Adjectives can give rise to about 18 different inflections. Verbs have 15 inflected forms, and also four participles. These participles each have 14–18 inflections of their own, and all the resulting 79 inflected forms may be further inflected to make them reflexive, giving a total of 158 inflections of the same stem. For various reasons, not all these inflections need occur on a given verbal stem, but most of them do.

Because of the preponderance of nouns in a dictionary, and the existence of words which are uninflected, or are best treated in the fully inflected form, it would be reasonable to estimate that a fully inflected dictionary would be about ten times the size of a stem-paradigm dictionary. Our earlier estimate of the size of the dictionary assumed a stem-paradigm organization, and also an advanced design of tape system. If a similar total reading time were to be achieved with ten times the quantity of dictionary data, the cost of tape transports would increase in proportion, and more elaborate logic in the transfer of data from tape to computer might be needed. There is therefore a strong case for some degree of condensation of the data. The system described below deals with the great majority of nouns and adjectives by means of one entry apiece, and with most verbs by one or two entries. Some other inflected words, for example pronouns, may need four or five items, but their total bulk is very small. The stem-paradigm entries are no bigger than they would be if they represented only one fully inflected word.

The example given above for the inflection of СТОЛ shows a single invariable stem, and 10 endings. This is the most regular type of inflection, and, fortunately, the most common in Russian, at least among nouns and adjectives. We must have the means for representing in our dictionary the less regular forms of inflection, and in this case we can sometimes choose two part-paradigms with different stems.

	Singular number		Plural number	
Nominative case	БРАТ	(brother)	БРАТЬЯ	(brothers)
Accusative case	БРАТА		БРАТЬЕВ	
Genitive case	БРАТА		БРАТЬЕВ	
Dative case	БРАТУ		БРАТЬЯМ	
Locative case	БРАТЕ		БРАТЬЯХИ	
Instrumental case	БРАТОМ		БРАТЬЯМ	

In this example, we have a regular set of endings for the singular on the stem БРАТ, and for the plural a regular set of endings on the stem БРАТЬ. Because endings like ЬЯ are not common enough to be included in our paradigm-coding, we are obliged to divide this declension into two halves, and therefore have two dictionary entries under the stems БРАТ- and БРАТЬ-.

Very irregular words might have to be treated in fully inflected form, with a separate entry for each inflected form. Such cases are rare in Russian, but small irregularities are not uncommon. The decision about what are the stems and what the endings is made for convenience, in each irregular example. There are some common types of stem-change, however, which we can treat by special methods. These are described later.

When this stem of the inflectional scheme of a word is examined, it is often possible to identify further suffixes which could be split off, and which are part of word-formation process, for example the verb УЧИТЬ (*to teach*) gives rise to the noun УЧИТЕЛЬ (*teacher*) in which Ь is the noun inflection and -ЕЛ- is a word-formation suffix. Words are also formed by prefixes. Both these processes could be used to condense the dictionary further, but their effect does not seem regular enough to give much advantage. We only use here the *syntactic* inflection (for case, number, gender, etc.) which behaves in a regular way for the great majority of Russian words.

The exact way in which a paradigm is coded in the dictionary entry will be described later. We will simply note here the functions that this coding must be able to perform. When a text word is found to agree with the stem part of a dictionary entry, removing this stem from it leaves the *affix*. From БРАТЬЯМ, for example, we would find the affix -ЯМ, and from СТОЛ the affix -* (null). It must then be possible to consult the paradigm information, and either (a) find that the affix is not compatible with it or (b) find that the affix is compatible, and also what case, number, gender, or other appropriate syntactic information it yields.

The paradigm-coding must be able to deal with four different sorts of entry:

(1) A full paradigm. This is the usual entry for an inflected word, and an example is СТОЛ.

(2) Part of a paradigm. An example is the singular part of the declension of БРАТ, or the plural part, БРАТЬЯ, etc. Many verbs have their inflected forms divided between two or three stems.

(3) Indeclinable words. For example, РАДИО (*radio*) which has the same form for all cases and numbers.

(4) Exceptional inflected forms. For example, ЛОШАДЬ (*horse*) has the instrumental plural form ЛОШАДЬМИ. Only five nouns can use the affix -ЬМИ so it is not one of the affixes we provide for in our paradigm-notation. A dictionary entry is needed for this fully-inflected form in which, in place of paradigm information, we have the syntactic data which would normally be the result of interpreting an affix (in this example: instrumental plural). Any exceptional parts of a paradigm will be dealt with in this way.

It is appropriate to state here a general principle governing the design of a dictionary organization. Wherever use is made of a rule or regularity in the language, provision must be made for escape, when the language does not conform. The provision for fully-inflected entries in a dictionary that was designed for stem-paradigm entries is an example of this 'escape principle'. No doubt the principle should be applied at all stages in mechanical translation schemes.

The use of a stem dictionary with and without a splitting process

The process for looking up words in a dictionary or magnetic tape begins with arranging the words in the same sequence as the words on the dictionary tape. (It is usual to employ alphabetic sequence, but any other process of ordering the words would also be suitable. We shall speak of an alphabetic sequence, for definiteness.)

A stem dictionary has its entries in alphabetic sequence of *stems*. The alphabetic sequence of *inflected words* is not exactly the same as the alphabetic sequence of their stems. An example of this from English is the verb 'to clear' with its inflections:

CLEAR
CLEARS
CLEARED
CLEARING

The noun CLEARANCE, with its inflection CLEARANCES, comes between CLEAR and CLEARS. Thus the stem sequence

CLEAR – CLEARANCE

differs from the sequence of the inflected forms

CLEARANCES – CLEARS

It would be expected that, because this inflection affects only the ends of words it would not change their sequence very much. The example given shows that related words will often give changes of sequence. When the stem is short, the number of dictionary entries by which a word shifts when it is inflected may be large.

To obtain the correct sequence of text words for dictionary look-up it would be necessary to identify their stems and arrange them according to these. A process for identifying the stem of a word is called a *splitting process* because it consists of splitting the word into a stem and an affix. An exact splitting process is not possible, for Russian, without employing a fairly large dictionary for the purpose. We shall describe an approximate splitting process which is reasonably simple, and show how this can be used to enable words to be looked up sequentially in a stem dictionary, with only a small addition to the size of the dictionary. It is not essential, however, to employ a splitting process, and in this section we shall briefly review some techniques in which the problem of correctly splitting an unknown word does not arise.

If we try to look up words in alphabetic sequence in a stem dictionary, we will occasionally find that the next word to be looked up has a stem that lies earlier than the last stem consulted. In this case, it is necessary to go back along the dictionary tape. Consider as an example the English word BRING. With no dictionary to help us we would have to consider two possible splits

BRING – *

and BR – ING

BR- comes earlier in the dictionary than BRING-. It is not possible to know that the stem of BRING has been passed until the next dictionary entry after BRING- has been reached. In the case of a word being looked up which is not in the dictionary, the search will always go as far as the dictionary stem equal to the whole word before it is abandoned.

It is necessary also to know the earliest point at which the stem of a word can be found in the dictionary, so that a complete search can be made. This corresponds to the maximum possible split, BR-ING in our example. Possible splits could be determined by looking up a table of affixes.

In the system described by Toma (1959) sequential searching through the dictionary continues until a particular text-word is not found. It is possible that the word has been missed by starting to look for it too late in the dictionary. Therefore the search goes back to a place in the dictionary determined by splitting off three letters from the text word. If this second search is still a failure when the place in the dictionary appropriate to the whole text word is again reached, the search for that word is abandoned.

Shifting the dictionary search back to an earlier position does not necessarily mean that the tape must be moved back, because a certain amount of the dictionary can be kept in the internal store of the computer at any one time. Provided the shift does not go further back than the stored data it does not involve reversing the tape.

An alternative approach is to look up the stems resulting from all reasonable splits. In the example, both the stems BRING- and BR- would go into the sorting process ready for look-up. Unfortunately, in Russian this would increase the amount of look-up by rather a large factor.

The decision between different methods of dictionary organization is very difficult. It depends in a most complicated way on statistics of the occurrence (a) in texts and (b) in the dictionary of all the different paradigms in conjunction with different endings of stems. Since the necessary data do not exist, the only method of deciding would be a comparative test. The effort to organize one dictionary system is considerable. Unless the first system fails, it is unlikely that a second system will be tried on the same dictionary data, with the same texts and the same type of computer.

We shall deal in detail with a dictionary organization in which splitting is attempted before looking up.

THE USE OF AN APPROXIMATE SPLITTING PROCESS

We owe to the Harvard Computation Laboratory M.T. Project the idea that an approximate splitting process can be used if it is applied both to text and dictionary (Oettinger 1958, 1960).

Given a word W, a splitting process defines two parts of this word: S′, the apparent stem, and E′ the apparent ending. We use the primes to distinguish these from the true stem S and ending E. We denote these two splits by the 'equations' W = SE = S′E′. If a fully inflected word W is split into S′E′ and entered in the dictionary under S′, it will be correctly found provided a text word is split in the same way before look up. Consider now a paradigm of *n* different endings $E_1, \ldots E_n$. The stem they are attached to is also important because the approximate split may cut into it. Let this be S. Each inflected form $SE_1, SE_2, \ldots SE_n$ must be entered in the dictionary in the split given by the approximate process. This can be described as $SE_i \rightarrow W_i \rightarrow S'_iE'_i$, i.e., the combination SE_i forms a word W_i which is split into apparent stem S′ and apparent ending E′.

Only with a split procedure which happens to work correctly for this paradigm would all the S'_i equal S. We have stated that no split procedure can always achieve this. If it is not achieved, there are still two possibilities:

(1) $S'_i = S'$ for all i (1 to *n*) but $S' \neq S$;

(2) condition (1) is not satisfied, i.e., the S'_i take more than one form.

Case (1) does not result in a correct split, but we shall call it a *consistent* split because it results always in the same apparent stem S′. It has the effect that the set of apparent endings E'_i does not agree with the paradigm E_i. With suitable organization, however, a consistent split can lead to a satisfactory single dictionary entry, under S′.

Case (2) must result in more than one dictionary entry. In practice it is very common to find two entries one of which is the correct stem S and serves for most of the paradigm.

Splitting is a technical trick for reducing the dictionary as far as possible to one entry per stem. It will have failed if too many entries behave as case (2) above. While it would be pleasing to devise splitting rules which usually gave the correct stem, it is only necessary to achieve consistent splitting.

There must, however, be a limitation on the extent of splitting, or the whole process would work with a split at the very start of the word, which is certainly consistent, on our definition. The limit comes about because of the appearance of *stem homographs*. With too much splitting, words of quite different original form will reduce to the same apparent stem. When a text word has to be looked up, it must be checked against all the entries having the same apparent stem. This can become a dictionary look-up problem in itself, with a number of text words to be checked against a number of dictionary entries, all these having the same apparent stem. The splitting process must not, therefore, introduce many stem homographs.

Description of a splitting process

It is important to clarify our use of a few technical terms. If a word is split in a way different from the standard stem-paradigm split, this is called a *false-split*. False splits are not necessarily undesirable, provided they are consistent. The part of a word split off will be called an *ending*. It may be regarded as made up from several parts. These will be called *affixes* in general, but the last one, which ends the word, may be referred to as an ending.

There is no unique solution to the problem of devising a splitting process, nor is there a good measure of the merit of any proposed solution. A compromise must be made between simplicity, consistency (in the sense defined above) and reduction in the size of the ending.

We will give a short description of the Russian inflection system, from our limited point of view, which will serve as a preparation for the splitting rules. In Table 1 we list the endings for nouns, adjectives (including short forms) and verbs, but we are not at present considering the participles explicitly. Some endings are absent here which may be known to the reader, for example the infinitive in - ЧЬ. Endings are rejected if they do not occur in a large class of words (example -ЬМИ) or if the stem to which they were attached does not serve for the remainder of the paradigm (example - ЧЬ). Our acceptance of paradigms was also connected with the technique for coding them, which will be described later. We call the affixes in Table 1 *first order affixes*. Any of the endings under VERB can be followed by the further reflexive ending СЬ (after a vowel) or СЯ (after a consonant). These will be called *zero order affixes*.

The participles form a special problem, if we wish to include them under the dictionary entry of their verb. They are formed by an affix added to the verbal stem, followed by an adjectival affix, and all this may be followed by the reflexive ending СЯ. It would be possible to make special dictionary entries for all participles (as adjectives) but we decided, for economy, to code them with their verbs. Splitting of three affixes in succession must therefore be allowed for. Table 2 shows the participle affix and the allowable adjectival affixes. Active and passive participles are listed separately because of rules of spelling which affect vowels after Ш and Щ, among other things. Short passive participles are also shown separately, because the participles in - НН- reduce to - Н- for these forms. We call perticiple affixes *second order affixes*.

A possible splitting procedure would consist of three steps:

(a) Split off -СЬ or -СЯ if it occurs at the end of the word (zero order affix).

(b) Split off the biggest first order affix from the remainder of the word.

(c) If the first order affix occurs in any of the lower lists in Table 2, split off the biggest second order affix in the corresponding upper list in Table 2.

In the procedure described, the association of second order affixes with certain first order affixes has been taken into account. For example - ННУ would not be split off, only У would be split in this case.

The question arises of making rules which limit the application of the splitting of zero order affixes. Splits in which these were found at step (a) could be disregarded unless the affix removed at step (b) appeared in the verb list or a second order affix was found. The correct association of either СЬ or СЯ could also be checked.

In making decisions concerning alternative splitting rules of this kind, we have

TABLE I

NOUN		ADJECTIVE		VERB	
•	Ь	ЫЙ	ИЙ	ИТЬ	ТЬ
	Й	ОГО	ЕГО		Ь
А	Я	ОМУ	ЕМУ		ТИ
У	Ю	ОМ	ЕМ	У	Ю
ОМ	ЕМ	ЫМ	ИМ	ИШЬ	ЕШЬ
О	Е	АЯ	ЯЯ	ИТ	ЕТ
Ы	И	УЮ	ЮЮ	ИМ	ЕМ
ОВ	ЕВ	ОЙ	ЕЙ	ИТЕ	ЕТЕ
АМ	ЯМ	ОЕ	ЕЕ	АТ	ЯТ
АМИ	ЯМИ	ЫЕ	ИЕ	УТ	ЮТ
АХ	ЯХ	ЫХ	ИХ	И	Й
ОЙ	ЕЙ	ЫМИ	ИМИ	ЙТЕ	ЬТЕ
	ЬЮ	•		А	Я
		А		ИВ	В
		О	Е	ИВШИ	ВШИ
		Ы	И		ШИ
				ИЛ	Л
				ИЛА	ЛА
				ИЛО	ЛО
				ИЛИ	ЛИ
					•

First order affixes.

used so-called 'end alphabetized' lists of Russian words. Normal alphabetic order is obtained by arranging according to the alphabetic sequence of first letters, and within groups of words having identical first letters, according to second letters, and so forth. In end alphabetization, the first arrangement is according to the last letters of the words, then within groups of words having identical last letters, according to the penultimate letters, and so forth. All the words in the end alphabetized list having a given ending, such as -ЕСЬ, can easily be found.

We found that very few Russian words give wrong splits which would be corrected by these improved rules for the reflexive. There are a few nouns in -ЕСЬ, -ОСЬ and -ЫСЬ, but a further complication is that -ЕТЕСЬ, -ЛОСЬ and -ЛЫСЬ are correct verbal endings. The extra complication is not worth-while.

We propose to show in the following section that the limitations implicit in Table 2 on the association of first and second order affixes cause inconsistent splitting, and that a simpler scheme, in which zero, first and second order affixes are split without any limitations gives greater consistency.

TABLE 2

ACTIVE PARTICIPLE AFFIXES	PASSIVE PARTICIPLE AFFIXES	SHORT PASSIVE PARTICIPLE AFFIXES
ВШ	НН	Н
Ш	ЕНН	ЕН
ИВШ	Т	Т
ЮЩ	ИТ	ИТ
УЩ	ИМ	ИМ
ЯЩ	ЕМ	ЕМ
АЩ	ОМ	ОМ
CORRESPONDING ADJECTIVAL AFFIXES	CORRESPONDING ADJECTIVAL AFFIXES	CORRESPONDING ADJECTIVAL AFFIXES
ИЙ	ЫЙ	•
ЕГО	ОГО	А
ЕМУ	ОМУ	О
ЕМ	ОМ	Ы
ИМ	ЫМ	
АЯ	АЯ	
УЮ	УЮ	
ЕЙ	ОЙ	
ЕЕ	ОЕ	
ИЕ	ЫЕ	
ИХ	ЫХ	
ИМИ	ЫМИ	

Participle affixes and their corresponding adjective affixes.

SIMPLIFICATION OF THE SPLITTING PROCESS

Many adjectives have a stem ending in -НН-, and are therefore not distinguishable from participles, for example ТРАДИЦИОННЫЙ (*traditional*). The long forms of the adjectives would be split wrongly, before the -НН-, but the split is consistent. In the example, the stem would appear as ТРАДИЦИО-. In the short form however, to keep the consistent stem it would be necessary to split the endings -НЕН, -ННО, -ННА, -ННЫ. There are also many adjectives with a stem ending in -Н-, which give rise to short forms in -Н (or -ЕН), -НА, -НО, -НЫ. The whole of these endings would be split off, because they are indistinguishable from short form participles. For consistency of splitting, it is therefore necessary to split off the adjectival endings -НЫЙ, etc. Adjectives with stems ending in -ЕН- and -ЕНН- give similar problems, which are solved by splitting off the adjectival endings -ЕНЫЙ, etc., and the short form endings -ЕНЕН, -ЕННО, -ЕННА

TABLE 3

PASSIVE PARTICIPLE AFFIXES	CORRESPONDING ADJECTIVAL AFFIXES
НН	ЫЙ
Н	ОГО
ЕНН	ОМУ
ЕН	ОМ
Т	ЫМ
ИТ	АЯ
ИМ	УЮ
ЕМ	ОЙ
ОМ	ОЕ
	ЫЕ
	ЫХ
	ЫМИ
	*
	А
	О
	Ы

and -ЕННЫ. All these requirements are met if the centre and right hand columns of Table 2 are merged, as shown in Table 3.

The special forms of 'second order affix' НЕН and ЕНЕН are not shown, because they need only be split if the first order affix is null. They will be discussed later.

We have now shown that in one instance, consistency of splitting can be improved by reducing the constraints linking first and second order affixes. This argument can be extended.

Consider the first order affixes *, А, О, Ы, ОМ and ОЙ. They occur in passive participles as first order affixes, therefore they allow the further splitting of second order affixes НН, Н, ЕНН, ЕН, Т, ИТ, ИМ, ЕМ or ОМ. But nouns can have these first order affixes, and if their stems end in Н, Т, ИМ, ЕМ or ОМ an apparent second order affix will be split as well.

To preserve consistency in nouns, which is important because of their preponderance in the dictionary, it would be better to allow the splitting of the passive participle affixes with any of the noun endings.

We must consider the possibility that only the hard endings of nouns should be associated also with second order affixes of the passive participle type. But unfortunately, the distinction of hard and soft endings fails with the ending Е, which is common to both.[1] To obtain consistent splitting of nouns, we must allow the splitting of any noun endings to be followed by the possible splitting of passive participle second order affixes.

We allow the splitting of ЕЙ to be followed by the splitting of active participle

[1] *Orthographic rules give apparently mixed declensions, but never in cases in which passive participle second order affixes occur.*

second order affixes. But ЕЙ is a noun ending, and this would result in the inconsistency

КАРАНДАШ | (*pencil*)
КАРАНДАШ | А
КАРАНДА | Ш | ЕЙ

(We show the two orders of splitting, where they occur, by separate vertical lines.) This inconsistency can be removed by allowing the splitting of any second order affix after any affix in the noun or adjective list.

The argument cannot be extended to verbs, because in the case in point (a stem which happens to end in a second order affix) the inconsistency cannot be avoided· Take as an example the verb ПЛАТИТЬ (*to pay*). On the principle that noun and adjective splitting must be kept consistent at the cost of verbs, because of their greater numbers, we have these splits for inflected forms of ПЛАТИТЬ:

ПЛА | Т | ИМ
ПЛА | Т | И
ПЛА | Т | Я

But the participles, as we intended, split

ПЛАТ | ЯЩ | ИЙ
ПЛАТ | ИМ | ЫЙ
ПЛАТ | ИВШ | ИЙ

This inconsistency seems inescapable. If the repeated removal of second order affixes were allowed, this would give greater overall consistency, but probably too many homographic entries. The trouble arises when the *stem* of a verb happens to end in a second order affix.

An examination of an end-alphabetized list of Russian words by Bielfeldt (1958) shows that verbs in -ТИТЬ and -НИТЬ are the commonest offenders. There are about 1,000 of these and, in all, about 1,500 verbs whose stems end in second order affixes. This is about 8 per cent of the verbs listed. Bielfeldt's list cannot be taken as representative of a technical or scientific glossary, but it probably gives a correct order of magnitude. The problem we are considering is a basic defect of the second order affix splitting scheme. It adds an extra dictionary entry to about 8 per cent of verbs. But the alternative, of having separate entries for all participles, would multiply the verbal entries by a factor much greater than 1.08.

The purely verbal first order affixes are given in Table 4. We shall now consider whether to make a rule that all these may not admit second order splits. Simplicity considerations would tend to make us avoid this rule, and have a free splitting of second order affix whatever occurred before. When inconsistency occurs, however, it will be found that dictionary look-up organization is simplified if as many words as possible are split to the lesser extent. This favours the restrictive rule.

TABLE 4

ИТЬ	ИТЕ	ЬТЕ	Л
ТЬ	ЕТЕ	ИВ	ИЛА
ТИ	АТ	В	ЛА
ИШЬ	ЯТ	ИВШИ	ИЛО
ЕШЬ	УТ	ВШИ	ЛО
ИТ	ЮТ	ШИ	ИЛИ
ЕТ	ЙТЕ	ИЛ	ЛИ

Verbal first order affixes.

Consider a noun like ДОЛОМИТ (*Dolomite*). Without the restriction, the splitting of three relevant forms is

ДОЛ | ОМ | ИТ
ДОЛОМ | ИТ | У
ДОЛ | ОМ | ИТЕ.

With the restriction it is

ДОЛОМ | | ИТ
ДОЛОМ | ИТ | У
ДОЛОМ | | ИТЕ

where a first order split not allowing a second is shown by | | . The restriction on a first order affix, preventing it from having a second order split, is useful if the first order affix applies only to verbs, and is identical with a second order affix or a second order affix followed by a first. A list of these follows:

ИТЬ, ТЬ, ТИ, ИТ, ИТЕ, ИВШИ, ВШИ, ШИ

It is not certain that all these can in fact occur, preceded by second order affixes, in actual nouns, but nouns ending in -ОМИТ -ИМИТ are common in technical literature. The splitting of other purely verbal affixes, for example -АТ, -ИШЬ, can cause inconsistency in nouns but this is made neither worse nor better by the restrictive rule. It turns out that application of the restriction costs nothing in size of programme and saves time, so we adopt it, for all purely verbal first order affixes.

The possibility was considered of splitting the affixes of the list immediately above in two stages, for example - | ИТ | Ь instead of - | | ИТЬ. This would reduce the list of first order affixes, but it created great programming difficulty in correctly identifying the affix that has been split, so it was not adopted.

Nouns with a stem ending in Л will generally give inconsistency in the splitting of -Л -ЛА -ЛО or -ЛИ. If the ending is -ИЛ, this is split in the same way. This problem is removed by

(a) adding Л, ИЛ to the second order affixes
(b) retaining Л, ИЛ, ЛА, ИЛА, etc., as first order affixes, in order to have correct identification, but forbidding them to be preceded by second order affixes.

We then obtain consistent splitting for

МЕТАЛ | | Л (*metal*)
МЕТАЛ | | ЛА
МЕТАЛ | Л | У

МЕТ | | ИЛ (*methyl*)
МЕТ | | ИЛА
МЕТ | ИЛ | У

ЗЕМ | Л | Я (*earth*)
ЗЕМ | | ЛИ
ЗЕМ | Л | Е

The *second order* affixes Л and ИЛ are special, in that they can never belong to the true ending, but are always part of the stem. They need no grammatical codes, therefore. The addition of Л and ИЛ to the second order affix list creates inconsistent splits in more verbs, specifically those in -ЛИТЬ. Bielfeldt's list gives 500 of these. But it gives about 1,500 nouns in -Л, -ЛА, -ЛО and -ЛЬ which will split inconsistently unless Л is included among the second order affixes.[2] The productive class of agent nouns in -ТЕЛЬ are among these, and account for about 500 in Bielfeldt's list.

A description of the splitting process which we have now developed is:

(a) Remove СЬ or СЯ from the end of the word, if these combinations occur
(b) Remove the largest affix in Table 1 from what is left
(c) If this is *not* an affix in Table 4, remove the largest affix in Table 5 from what is left.

Our splitting procedure has been designed for consistency rather than grammatical correctness. Wrong splitting is not particularly disadvantageous if it is consistent. But there is one kind of splitting error which would cause great difficulty in affix interpretation. This would arise if a participle (which has both a first and a second order true affix) had an *apparent* first order affix split off which penetrated into its second order true affix.

Fortunately the adjectival affixes, which can occur in these situations, never give rise to false splits, and the short form affixes *, А, О, Ы, never give rise to false splits when they are preceded by second order affixes. In the case of the short form participles in -Т, the first order split can remove АТ, ЯТ, УТ, ЮТ, ЕТ or ИТ if any of these occur. We must therefore note the possibility that - | Т may be the correct split. This is in practice dealt with by adding | | Т to the list of first order affixes not allowing a second order split.

The affix -АТ occurs only as a variant of ЯТ occurring after the consonants Ж, Ч, Ш or Щ because of spelling rules. We therefore need only split -АТ if the pre-

[2] *The foreign borrowings in -ль usually have nominative plural in Я, and are unaffected by these inconsistencies.*

TABLE 5

НН	ИМ	ЮЩ
Н	ЕМ	УЩ
ЕНН	ОМ	ЯЩ
ЕН	ИВШ	АЩ
Т	ВШ	ИЛ
ИТ	Ш	Л

Second order affixes.

ceding letter is one of these four. Our method of splitting allows this to be done, and we represent the rule by

Ж | | АТ, Ч | | АТ, Ш | | АТ, and Щ | | АТ

This notation means that ЖАТ, etc., must be recognized, but only -АТ is split, and no second order split is allowed.

SPECIAL RULES TO DEAL WITH PRODUCTIVE NOUN FORMATIONS

There cannot be a general rule which would indicate, for example that АТОМ was not the instrumental singular of АТ. Therefore exact, or even consistent splitting is impossible to obtain. There is some value, however, in examining the process of word formation in Russian to see if any very productive methods of suffixial word formation give rise to inconsistent splits. If they do, additional rules may be made which avoid the inconsistency without spoiling other words not formed by this process. All of the workable examples of this seem to have been discovered by the Harvard group.

We have taken our data concerning noun formation from Unbegaun (1957), and we give details here of the cases in which extra rules can result in correct splitting.

Nouns of action formed from verbs by the addition of -НИЕ or -ТИЕ to the stem would give rise to inconsistent splits in the nominative singular -НИЕ or -ТИЕ and the genitive plural -НИЙ or -ТИЙ. Examples are РАСПИСАНИЕ (*timetable*) and ВЗЯТИЕ (*capture*). If we examine one or two more letters than usual at the first order affix stage we can make special provision to split the following letter combinations as shown

-АНИ|Е	-АНИ|Й
-ЯНИ|Е	-ЯНИ|Й
-ЕНИ|Е	-ЕНИ|Й
-ТИ|Е	-ТИ|Й

This corrects the nouns without spoiling adjectives in -ИЙ, because they do not have this combination of letters in their stem.

Unfortunately, nouns of action formed by -НЬЕ and -ТЬЕ give rise to a false split in the dative singular which cannot be corrected without spoiling the consistency of split in nouns in -НЬ and -ТЬ.

Nouns of action formed by the addition of -ИЗАЦИЯ give rise to a false split in the genitive plural -ИЗАЦИЙ. This can be prevented by a special rule for first order affix splitting: -ЦИ|Й. This allows for all those nouns in which an English *-tion* is rendered in Russian by -ЦИЯ. Examples are АВИАЦИЯ (*aviation*) and ИОНИЗАЦИЯ (*ionization*).

Abstract nouns of quality can be produced by the suffix -СТВО, and they give rise to a false split in the genitive plural -СТ|В. An example is ЗВЕРСТВО (*ferocity*). This combination of letters does not arise in the past gerund for which the В ending is designed, so we make the special rule -СТВ|* for this combination when splitting first order affixes.

The Harvard split procedure takes account of the suffix -СТВИЕ, an example of which is ВЕДСТВИЕ (*hardship*). Its nominative singular and genitive plural give trouble by splitting before the И, although the word is a noun with И on its stem. The special rules -ВИ|Е and -ВИ|Й deal with this, and no adjectives are spoilt by them.

There are four quite productive affixes for noun formation which our split procedure deals with consistently. Agent nouns can be formed by the suffixes -ТЕЛЬ and -ЛА. Examples are МЕЧТАТЕЛЬ (*dreamer*), МЕНЯЛА (*money changer*) and ГРОМИЛА (*burglar*). All these give rise to forms which must split like the past tense of a verb. But the decision to include Л and ИЛ among the second order affixes was made with these in mind, and they now split consistently in front of the Л or ИЛ. Abstract nouns can be formed from adjectives by the suffixes -ОСТЬ and -ЕСТЬ. The former is described by Ungegaun as 'the most frequent suffix'. The nominative of singular forms must split like infinitives of verbs. Examples are ГОРДОС | ТЬ (*pride*) and СВЕЖЕС | ТЬ (*freshness*). The inclusion of Т among the second order affixes ensures a consistent split in these cases.

The treatment of false splits in the dictionary

We have now determined a set of permissible affixes which will be split off. This includes all the affixes which are required by our paradigms, and any parts of actual paradigms which do not conform are treated as fully inflected entries. It follows that any false split of an inflected form contained in a stem-paradigm dictionary entry must have more letters in the affix and less in the stem than it should. That is false splits are oversplit. (Fully inflected dictionary entries are also split, but we are not concerned with accuracy of splitting for these, because there is no consistency problem.)

Let us consider for definiteness a consistent false split, the word УЧИТЕЛЬ (*teacher*). The stem is УЧИТЕЛ- and the endings Ь, Я, Ю, Е, ЕМ, etc. The apparent stem is УЧИТЕ- and the apparent endings ЛЬ, ЛЯ, ЛЮ, ЛЕ, ЛЕМ, etc. Both the dictionary entry and the split text word will be arranged in alphabetic sequence according to the apparent stem УЧИТЕ-. There is probably only one entry with apparent stem УЧИТЕ-, and this is found by the usual method of search. Before accepting the stem, however, the comparison of text word with dictionary entry should go as far as УИЧТЕЛ. We can express this by saying that the dictionary entry has a *split stem* УЧИТЕ|Л and that words are sorted on the part to the left of the split but compared with the whole entry. When the split text word УЧИТЕ|ЛЮ has been found to match with УЧИТЕ|Л the remainder -Ю must next be examined to determine whether it agrees with the paradigm of this entry, and if it does, to determine what -Ю signifies. The result would be "Yes", and "Dative Singular". This affix interpretation process is the subject of a later section.

It is possible to store the Russian word of entry УЧИТЕ|Л in precisely this form, the symbol | having a code combination of its own, and occupying the space of one letter. If | is made of earlier position in the alphabetic sequence than any Cyrillic character, sorting a set of such words would be equivalent to sorting on the apparent stem (УЧИТЕ-) followed by sorting groups of homographic apparent stems on the remaining letters. In general, this will put the true stems (У ЧИТЕЛ-) into sequence, but it may not, because of inflection. Provision must be made for *possibly* matching any text word having a given apparent stem against any dictionary entry with this stem. An actual example with a great degree of homography will be given later.

Let us now consider an inconsistent split, in particular the word ЗАЛИВ (*bay*). This gives rise to two entries, one under the unsplit stem ЗАЛИВ|*, which covers nearly all the inflected forms, and one for the nominative singular, which splits as ЗА|Л|ИВ. It always happens that the stems which are over-split occur earlier in the dictionary. We propose therefore to store in this place (ЗА|ЛИВ) only a small cross-reference entry which enables the programme, by examining the text word, to reconstruct the correct split. A cross reference entry is only one eighth of the size of a full dictionary entry. It contains essentially only a split Russian word, and the information about the correct split. We shall call this cross-reference entry a *short entry*.

The short entry may not always be derived from inflection with a null ending, nor need the entry to which it leads be a correct split. It can be thought of as a method for correcting over-split words to refer them to the least split form of the stem, under which the full dictionary entry is stored. But probably the great majority of cases will be like ЗАЛИВ, the null ending giving a false split and the remainder a correct split.

The effect of a short entry is to generate a new split text word for look up at a *later* point in the dictionary. A small store must be reserved for these items, and they must be kept in alphabetic sequence, each newcomer being merged into sequence. In this way a minimum of testing is needed to bring these cross references out at the correct point in the dictionary. Except for very short stems, they will not be held in the store for long. If a few frequent words give trouble by filling the small store provided, this can be avoided by putting full dictionary items in place of these short entries. An alternative to this organization is described in the next section of the paper.

We shall now give an example of an exceptionally complex set of related short and long entries. It shows all the complexity that is possible, and is a useful test for the scheme. The corrected split of a short entry is shown by the second of the two vertical bars, the first bar showing the split expected from the splitting routine.

НАЛЕТ (*raid, thin coating*) gives rise to the following entries:
Short entry НАЛ|Е|Т from НАЛ | | ЕТ and НАЛ | | ЕТЕ
Full entry НАЛЕ|Т from all other inflections

НАЛИМ (*eel-pout*) gives the following entries:
Short entry НА|Л|ИМ from НА|Л|ИМ
Full entry НАЛ|ИМ from all other inflections

НАЛИВ (*sap or juice of fruit*) gives the following entries:
Short entry НАЛ|ИВ| from НАЛ | | ИВ
Full entry НАЛИВ|* from all other inflections

НАЛОЙ (*lectern*, in popular speech) gives the following entries:
Short entry НА|ЛО| from НА|Л|ОЙ НА|Л|ОЕ
Full entry НАЛО|* from all other inflections

НАЛИТЬ (*to pour out,* perfective) gives the following entries:
Short entry НА|Л| from НА|Л|Ю and other inflections
Full entry НАЛ|* from НАЛ | | ИТЬ and other inflections

The set of entries, in sequence, is:

1. Short НА|Л|
2. Short НА|ЛО|
3. Full НАЛ|*
4. Short НАЛ|Е|Т
5. Short НАЛ|ИВ|
6. Full НАЛ|ИМ
7. Full НАЛЕ|Т
8. Full НАЛИВ|*
9. Full НАЛО|*

Note that the short entry 1, makes the more restricted entry НА|Л|ИМ unnecessary, but there may be organizational reasons for leaving the latter entry in the dictionary. The word НАЛИВ would be split НАЛ|ИВ and would match the full entry НАЛ| and the short entry НАЛ|ИВ. The latter would generate the word НАЛИВ|* which would match against НАЛИВ|*. These correspond to two meanings, the nominative singular of НАЛИВ and the past gerund of НАЛИТЬ.

If the two words НАЛ/ЕТ and НАЛ|ИВ were being looked up, they would be in the order given, but НАЛ|ЕТ must be tried against entries 3 and 4, while НАЛ|ИВ must be tried against 3 and 5. This demonstrates that each word with the apparent stem НАЛ- may have to be tried against any entry, short and full, which has this apparent stem. The degree of homography shown by this example is, fortunately, rare.

An alternative method of storing short entries

The proposal described above, for short entries, assumes them to be mixed with long entries in one alphabetic sequence. The look up takes place in one pass of this combined tape, at the cost of keeping a running store of newly split words which are in correct sequence.

A simpler alternative is available, for which programming is easier, but the process may be slightly slower. It is to keep a separate dictionary of short entries. The text words, having been arranged in alphabetic sequence of apparent stem, are first tested against the short entry dictionary. A few will generate new splits; which will be stored as a separate tape. These may be slightly out of sequence, and a sorting (arranging) programme, designed for such a near-correct sequence, will be used to re-arrange them. Then a dictionary look up will be made in the dictionary of full entries, two look-up tapes being run at once, one containing *all* the original text words and the other containing the newly generated splits.

Treatment of mobile vowels by short entries

In Russian nouns, short form adjectives, short form participles and the past tense of a few verbs, the paradigm includes the null ending. In some of these cases a vowel, generally O or E is inserted before the last letter of the stem. For example the normal stem for the word *father* is ОТЦ- but the null-ending form is ОТЕЦ, with an inserted E. This E is called a mobile vowel.

There is a second variety of mobile vowel in which the vowel in the null ending case replaces a soft sign which is present in all other cases ; as an example ЛЕВ, ЛЬВА (*lion*). After a vowel the soft sign becomes Й; as an example ЗАЕМ, ЗАЙМА.

The mobile vowel occurs with null endings, and it can also occur with the ending Ь ; as an example ДЕНЬ, ДНЯ (*day*).

A mobile vowel usually gives rise to a different apparent stem from the remainder of the inflections. This can, however, be treated by a cross reference entry, no bigger than that required for a false split. In order to ensure that the cross-reference entry comes earlier than the paradigm to which it refers, we make our alphabetic order different from usual, and place the vowels first. Thus the text word ОТЕЦ is matched with a short cross reference item under ОТЕЦ|* which constructs the fictitious word ОТЦ|*. This is held in the store for cross references until the ОТЦ|* item in the dictionary arrives. The procedure has a loophole, in that ОТЦ might be a Russian word of different meaning. We have not found examples of this phenomenon, but it could be safeguarded by a mark on the cross reference and a mark in the dictionary item to say that only such marked cross references are accepted with * or Ь endings. We shall now give the rules used for treating mobile vowels.

Mobile vowel short entries are of two types: delete (i.e., delete vowel) and substitute (i.e., substitute Ь for vowel). The only coding needed in short entries is that which defines a false split entry or one of these two mobile vowel types. The effect of a mobile vowel entry on a word which matches it is to transform that word. Therefore the splitting of the original word is no longer valid, and the mobile vowel rules given below define the new split.

Rules for entries of the 'delete' type

(Rules apply to the whole text-word, previous splitting being ignored, but the word must match the entry exactly. Rules for the new split are given.)

(1) If the final letter is Ь:
Delete the letter before the letter before the Ь. Split off the Ь, and if the previous letter is Т, or Л, split this as well.

(2) If the final letter is Й:
Delete the last two letters and add Ь. Split after the Ь.

(3) In other cases:
Delete the letter before the last letter. Split after the last letter, unless the word now ends in Т, Л, Н, НН or ЕНН, which should be split.

Rules for entries of the 'substitute' type

(1) If the letter before the letter before the last is a consonant:
Replace the letter before the last by Ь. Split after the last letter unless it is Т or Л, which should be split.

(2) If the letter before the letter before the last is a vowel:
Replace the letter before the last by Й. Split after the last letter, unless it is Т or Л, which should be split.

These rules deal with all the common examples of mobile vowels, for example the short masculine forms of adjectives in -НЕН or -ЕНЕН are reduced to - | НН and - | ЕНН and the short masculine forms of past passive participles and adjectives in -ЕН are reduced to - | Н. They also deal with unusual cases like ЗАЯЦ (*hare*) which is reduced to ЗАЙЦ | *. Some peculiarities of noun declension are dealt with, for example:

РУЧ|ЕЙ	(*stream*, n.s.) becomes РУЧЬ|*
ГОС|Т|ИЙ	(*visitor*, g.p.) becomes ГОСТЬ|*

The resultant forms are then consistent with the remainder of the paradigm, for example РУЧЬ|Я g.s. and ГОСТЬ|Я n.s. In all cases, the rules given, with the indication 'delete' or 'substitute' suffice to restore the correct stem. Provision is made for splitting a second order affix from this stem if necessary.

The endings НЕН and ЕНЕН have not been included in our lists of affixes, therefore they would split as -Н | ЕН and -ЕН | ЕН. This split, however, is not used in the mobile vowel rules, which generate the endings - | НН and - | ЕНН. The newly split words would then be earlier in alphabetic sequence than the originals. This is not acceptable if short and long entries are stored in one dictionary, but if they are stored separately it does not matter.

If long and short entires are stored together, it is necessary to give the splits

- |НЕН and - |ЕНЕН. We propose therefore to add the affixes - | | НЕН and - | | ЕНЕН to the first order list, and not permit them second order splits. This is better than having them in the second order list because it reduces the amount of splitting in the few words like

ПРИМЕН | | ИТЬ
short past passive { ПРИМ | | ЕНЕН (inconsistent, in any case)
participle forms { ПРИМЕН | ЕН | А

Having ЕНЕН as a first order affix, instead of second order, makes the maximum number of letters in the ending 6, apart from the reflexive affix. (The only four letter first order affix | | ИВШИ does not allow a second order split.) This maximum is attained in such endings as - | ЕНН | ОГО.

The programming of the split procedure, with affix identification

The first stage of splitting is the removal of СЬ or СЯ from the stem. This is very simple.

In the next two stages, the final letters of the word must, in effect, be checked against the lists of affixes. This could be done by storing the list of allowed affixes and testing them in turn until either one is found to match, or the end of the list is reached. This may require as many tests as there are affixes in the list, which is about 80 for the first order of splitting.

The first order affix list is long, but all the affixes end in one of the 15 letters А, Я, Ы, И, О, Е, У, Ю, Ь, Й, Л, В, Х, М and Т. A test of these fifteen would eliminate most words with a null first order split. If one of these occurs, there is a maximum of 6 penultimate letters which could then lead to an affix of two or more letters. For example И, leads only to ШИ, ЛИ, ТИ and МИ (of these, only МИ is not itself an affix, but it leads to АМИ, ЯМИ, ЫМИ and ИМИ).

Letter-by-letter tests of this kind result in an average of about 10 tests for first order splitting, in place of the average of 40 which would be needed if a single list were employed.

The programming of these tests could be done by writing instructions for each separate test. Undoubtedly the quickest programme could be made in this way. But for economy, we prefer to retain lists of affixes, and a common programme which refers to these for guidance on what tests to make next. Full affixes need not be listed. For example if И has been identified in the final letter, and this leads us to a list of penultimate letters, only the Ш, Л, Т and М need be stored, because the final И is already determined.

To illustrate the nature of the list we shall invent a very simple set of first and second affixes, to avoid the bulk of the real set. These affixes will, it is hoped, show all the features of the system. The splitting rules are summed up in Table 6.

The table employed in a programme for splitting according to this simplified scheme is Table 7.

TABLE 6

First order affixes

*, Ы, О, | | ЛО, | | ИЛО, ОГО, ЕГО, Е, ИЕ, ВИ | | Е, АХ, ЯХ, ЫХ,
| | Т, | | ЯТ, Ш||АТ, Щ||АТ

Second order affixes

*, Л, Н, НН, ЯЩ, АЩ

Example of splitting rules, for illustration only.

TABLE 7

(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
1	Ы	20	1	S	11	Ы	20	2	E,S	21	Н	23	1	
2	О	6	1		12	Я	*	2		22	Щ	24	1	B
3	Е	8	1		13	А	18	1	B	23	Н	*	2	
4	Х	9	0		14	И	*	3	B	24	Я	*	2	
5	Т	12	1	E	15	О	20	3	S	25	А	*	2	B
6	Л	14	2		16	Е	20	3	E,S					
7	Г	15	1	E	17	В	*	1	E					
8	И	17	2	E	18	Ш	*	2						
9	А	20	2	S	19	Щ	*	2	B					
10	Я	20	2	S	20	Л	*	1						

Example of splitting table, for illustration only.

The five columns of the table are:

(1) Serial number of entry.
(2) letter to be tested against *current* letter of the Russian word.
(3) if the test is successful (i.e., a match) this gives the serial number of the next entry of the table to be used. If it is marked * the split procedure is completed.
(4) the number of letters to be split off the word if the test is successful.
(5) S — The second order affix is now to be considered (Col. 3 = 20).
E — End of the present list, look for second order affix.
B — End of the present list, if it does not match, the split procedure is completed.

The procedure followed by the programme for each step of the process will now be described. At the start of the process, the *current letter* of the Russian word is its last letter. As the tests proceed, the next letter to the left becomes the current letter, then the next, and so forth. This is referred to as 'shift left'.

Compare the current letter with the letter in column (2). If it matches, record the split number from column (4), record S if it is given in column (5), shift left, and go to the table entry given in column (3). If column (3) shows * the process is complete. If it does not match, and if column (5) does not show B or E, move to the next table entry. If it does not match, and if column (5) shows B the process is complete. If column (5) shows E, adjust the current letter position so that it is the letter preceding the split indicated by the recorded split number, record S and go to table entry 20.

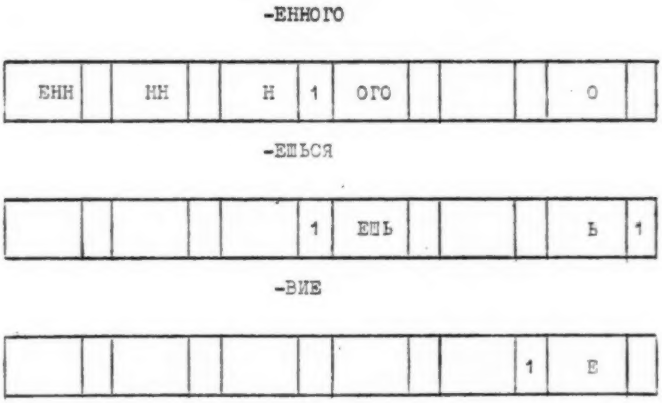
The split number, when recorded, overwrites any previously recorded split number. After S is recorded, however, the split number is recorded in a different place. Thus, split numbers for the first and second order splits are maintained separately. When the end of a list is reached without any match being found, the correct split to be applied is the last one that was found. This may be of the type that leads to second order affix or of the type which does not. The letters E or B indicate this. The output of the process we have described would be two numbers, giving the numbers of letters in the first order split, and in the second order split if this occurred.

At the stage of affix interpretation, we need to know not only how many letters have been split but also what letters they are. The letters themselves can be regarded as a coded form of 'affix identifier' but as such they are very inefficient. The 'code' of the affix ИВШИ contains 24 bits. As the next section of this paper shows, we require a reference to a table of about 100 entries. For this purpose a 7 bit number is associated with each affix and is called its 'affix identifier'.

Given the affixes, to find the identifiers is a job comparable with the affix splitting procedure. We propose a better solution, which is to record the affix identifiers in the table used for splitting. They can then be recorded when the split is made, and one process only is needed. The affix identification data must then be carried along with the text word through the process of dictionary look-up to the affix interpretation process.

When a short entry is employed, the amount of the split will be reduced. When the dictionary stem is a split one, the true affix will be less than the one given by the split procedure. It is necessary therefore, to carry along the affix identification data not only for the maximum split, but also for the lesser splits, because these may be the true ones. We intend to have a possible 6 affix identifiers (making up one machine word) corresponding to the six letters that may be split off, after the reflexive affix has been removed.

In Fig. 1 we show a possible format for the affix identifier word, with three examples. The 7 bit fields are separated by one bit fields which are used to indicate the limits of the first and second order affixes. The final single bit field indicates the presence on the original word of the ending СЬ or СЯ. This affix identifier word is formed during the splitting process. When a letter matches a table entry, not only is the split number recorded, but the affix identifier is extracted from the table and placed in the field indicated by the split number. If the split number has not increased (as in going from О to ГО) this need not be done. If the split number decreases (as in going from ИЕ to ВИЕ) some affix identifiers may have to be removed. The one bit fields are filled at the stage of entering the second order affix splitting procedure, and at the stage of leaving the splitting process.



-ЕННОГО

ЕНН		НН		Н	1	ОГО				О	

-ЕШЬСЯ

					1	ЕШЬ				Ь	1

-ВИЕ

									1	Е	

Fig. 1. The affix identifier word. The affix identification numbers are represented by the affixes themselves.

If the reflexive affix is shown, and either a short entry or matching with a dictionary stem demands a change of splitting that leaves only one letter on the ending, a special routine replaces the whole affix identifier word by the data appropriate to Я or Ь. It chooses between these by looking at the last letter of the Russian word.

In the next section we shall describe how the affix identifier word is used during affix interpretation. It is important to remember that at this stage the correct identification of first and second order affixes is impossible. For example, a word ending in -ЯТ could either have the first order affix ЯТ or the second order affix -T with an ending *. It could, alternatively, be a stem in -ЯТ with ending *.

The single bit fields in the format shown in Fig. 1 are merely a simple way of showing the division into apparent first and second order affixes. They could be replaced, if the programme demanded, by two fields giving the number of letters in these two affixes, i.e., the split numbers.

Paradigm coding and affix interpretation

We remarked earlier that a full entry in our dictionary must in general contain some coding for the paradigm, i.e., the set of endings that can be attached to the stem and the grammatical significance of each. (Entries for indeclinable words and for words with exceptional endings will not have this paradigm coding.) We shall now describe this coding, and the process by which a given ending is made to yield its grammatical significance. Full details of this process would be too elaborate for this paper; only the principles will be given.

The method of coding has been designed to facilitate the affix-interpretation process. Fig. 2 shows four formats for the 48 bit word employed in the computer. We shall use the noun format as an example. There is a nominative singular field of

7 bits, and these are associated with the 7 endings we recognize in our paradigms for this case and number. Each case and number has its field, with a bit position for each possible ending. The accusative case is an exception, but we will not consider this.

A paradigm for a noun is made up in general by selecting one ending from each field. It is coded by inserting in a 48 bit computer word a binary digit 1 in the appropriate place in each field. For example a straightforward masculine declension (ignoring accusative) is coded by inserting 1 in the binary digit positions 1, 11, 15, 19, 21; 26 30, 36, 38, 40. If a noun has alternative endings (for example, A or ЬI is the nominative plural) both may be included (digits 26 and 27 in our example). To represent only part of a paradigm one merely inserts the ones required and leaves out those for cases and numbers not represented.

These 'paradigm indicators' are compiled by the computer from data originally written on a suitable form. This form sets out the possibilities and guiding rules for a particular type of declension so that the linguistic choices to be made are reduced in number. Details of dictionary compilation are not dealt with in this paper.

The format for paradigm indicators differs, of course, for nouns, adjectives and verbs. We also have provided two formats for verbs, but only because 48 bits were not enough to accommodate all the endings. The two verb formats are identical in their location of grammatical fields. It will be noted that they do not specify the inflectional pattern of the adjectival endings of participles; they specify their second order affixes only.

Suppose now that we have an affix -У to interpret for a paradigm word in the noun format of Fig. 2. The affix У appears in the digits 8 and 15. If we have a word with only these digits as ones, and we collate it with the paradigm word (logical operation 'and') the result is a digit in the fields corresponding to the cases and numbers allowed by this paradigm for the affix У. For example, in the straightforward masculine declension the result would be digit 15 only, which is in the Dative Singular field. The word giving all the occurrences of У in the noun format is called the *noun role indicator* of У.

The role indicators of all the affixes we use are stored in a table of about 100 entries. To extract the appropriate role indicator we use the affix identifier obtained during splitting. This is the address in the table of the first of the set of role indicators for the given affix. For example У as an affix can serve in nouns, И verbs and E verbs, therefore in principle 3 role indicators are needed. If the affix identifier of У were 18, say, the three role indicators would be in lines 18, 19 and 20 of the table, and line 20 would be specially marked as the last of the set of 3. If У were being matched against, say, an И verb paradigm, the programme would search from 18 to 20 for the И verb role indicator.

Actually, the И and E verb role indicator for the affix У are identical. There are 10 affixes with this property У, Ю, И, Й, Ь, ЙТЕ, ЬТЕ, А, Я and ЕНН. Space is saved by marking their verbal role indicators in a certain digit to indicate their double usefulness.

NOUN

N.S.							A.S.			G.S.				D.S.				
*	Й	Ь	А	Я	О	Е	У	Ю	Ь	А	Я	Ы	И	У	Ю	Е	И	
1				5					10					15			18	

	L.S.		I.S.					N.P.				G.P.						
	Е	И	ОМ	ЕМ	ЕЙ	ОЙ	ЬЮ	Ы	А	И	Я	ОВ	ЕВ	ЕЙ	*	Ь	Й	
	19	20					25					30					35	

	D.P.		L.P.		I.P.								
	АМ	ЯМ	АХ	ЯХ	АМИ	ЯМИ							
	36				40						45		48

ADJECTIVE

MASCULINE											FEMININE						
N			G		D		L		I		N		A		GDLI		
ЫЙ	ИЙ	ОЙ	ОГО	ЕГО	ОМУ	ЕМУ	ОМ	ЕМ	ЫМ	ИМ	АЯ	ЯЯ	УЮ	ЮЮ	ОЙ	ЕЙ	
1				5					10					15		17	

	NEUTER										PLURAL				
	NA		G		D		L		I		N		G		
	ОЕ	ЕЕ	ОГО	ЕГО	ОМУ	ЕМУ	ОМ	ЕМ	ЫМ	ИМ	ЫЕ	ИЕ	ЫХ	ИХ	
	18		20					25					30	31	

	PLURAL						SHORT										
	D		L		I		M	F	Neut.		Plur.						
	ЫМ	ИМ	ЫХ	ИХ	ЫМИ	ИМИ	*	А	О	Е	Ы	И					
	32			35					40					45			48

Fig. 2(a). Formats for paradigm indicators.

И VERB

Infinitive			Present								Imperative						
			1		2	3	1	2	3		Present			Past			
ИТЬ			У	Ю	ИШЬ	ИТ	ИМ	ИТЕ	АТ	ЯТ	И	Й	Ь	ИТЕ	ЙТЕ	ЬТЕ	
1				5					10					15		17	

	Gerunds				Past						Active Participles				
	Present		Past		M		F	N	P		Present		Past		
	А	Я	ИВ		ИЛ		ИЛА	ИЛО	ИАИ		АЩ	ЯЩ	ИВШ		
	18		20					25						30	

	Passive Participles																	
	Present		Past				(short)											
	ИМ			ЕНН	ИТ		ЕН											
	31				35					40					45			48

E VERB

Infinitive			Present								Imperative						
			1		2	3	1	2	3		Present			Past			
ТЬ	ТИ	Ь	У	Ю	ЕШЬ	ЕТ	ЕМ	ЕТЕ	УТ	ЮТ	И	Й	Ь	ИТЕ	ЙТЕ	ЬТЕ	
1				5					10					15		17	

	Gerunds				Past					Active Participles				
	Present		Past		M		F	N	P	Present		Past		
	А	Я	В	ШИ	Л	*	ЛА	ЛО	ЛИ	УЩ	ЮЩ	ВШ	Ш	
	18		20					25					30	

	Passive Participles																	
	Present		Past				(short)											
	ЕМ	ОМ	НН	ЕНН	Т	Н	ЕН											
	31				35					40					45			48

Fig. 2(b). Formats for paradigm indicators.

The process of identifying a given affix is therefore

(a) From the affix identifier, find the place in the role indicator table
(b) Find the role indicator, by searching a small list of them (maximum of three).
(c) Collate with the paradigm indicators (logical 'and' operation)
(d) Condense the result of (c).

Step (d) refers to the fact that, taking nouns as an example, we need only a 12 bit output, each bit giving a case and number. Dative Singular, in the output of step (c) may be any of the digits 15, 16, 17 or 18. After condensation it would be a single digit. The exact form of affix interpretation output depends on the requirements of the syntax routine which will use it. It is probably desirable for noun and adjective outputs to be in the same form.

We have described affix interpretation only in the simplest case. Let us now consider the possibility in more detail. During splitting, an affix identifier word has been built up, with the format shown in Fig. 1. This indicates, among other things, whether two affixes have been found, or only one. Take as an example the word ДОЛОМИТУ, with the dictionary entry ДОЛОМИТ (*Dolomite*), and consider the tests needed to establish that the entry is compatible with the text word. The text word is split ДОЛОМ | ИТ | У and the dictionary entry is split ДОЛОМ | ИТ. In the sequence of alphabetically sorted text words, the given word is in the place determined by the letters ДОЛОМ. The dictionary entry is also alphabetically sorted under ДОЛОМ. The first stage of testing is therefore carried out when an entry with the apparent stem ДОЛОМ is found. At this stage, the parts of the words to the left of the split must agree. If there is more than one ДОЛОМ- entry, all such entries must be subjected to the remainder of the tests.

The next test is to extend the comparison to all the letters of the dictionary stem ДОЛОМ | ИТ. At the same time, the split of the text word is altered to match it, making ДОЛОМИТ | У. The affix which is now relevant is the first order affix -У. (If the affix identifier word (Fig. 1) did not have an entry in the single letter affix slot, the word and entry would be incompatible.) The affix identifier for У is then used in the affix-interpretation programme described earlier.

In our example, the two affixes originally split off have been reduced to one. There may be one or two affixes left at this stage. The affix to be treated first is the left most one, the one adjacent to the stem. There may be no affix left, which would happen if the dictionary stem and text words were of equal length. The procedure varies slightly in these three cases.

No Affix

Generate the role indicator for * and treat as one affix, but part (a) of the procedure only.

One Affix

(a) Noun or adjective paradigm—interpret the affix as described above. Verb paradigm—interpret as described above, but deleting the fields concerned with participles.

(b) Verb paradigm—additional step; generate the affix * to the right of the given affix and treat as two affixes.

Two Affixes

Noun or adjective paradigm—incompatible.

Verb paradigm—interpret the left most affix as described above, but deleting the fields not concerned with participles.

If it is compatible, identify the type of participle (active-passive) and take from the appropriate adjectival paradigm. Take the adjective role indicator of the right most affix and interpret as described above. The output is in two sections, one adjectival, and the other identifying the four types of participle.

Some short cuts are possible in the computer programme based on the above scheme. These enable incompatibilities to be recognised at an earlier stage.

Conclusions

We have described the organization of a stem dictionary on magnetic tape with the following features:

A single pass of the dictionary tape is sufficient to process a block of text words, without any reversals. In an alternative version of the scheme, a small sub-dictionary is referred to in advance, and this generates a small set of additional text-words which must be arranged in sequence before proceeding to the main dictionary.

The space occupied is very little greater than that needed for a perfect stem dictionary. Mobile vowels are treated with very little extra space or complication of programme.

These features can be combined with a very simple method of representing paradigms and of interpreting affixes.

The work described above has been carried out as part of the research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory.

References

Toma, P. (1959). Serna System. Part I Morphology, Machine Translation Programming Paper 1, Georgetown University, Washington, D.C.

Oettinger, A. G., Faust, W., Giuliano, V. E., Magassy, K. and Matejka, L. (1958). Linguistic and machine methods for compiling and updating the Harvard Automatic Dictionary. Preprints of papers for the International Conference on Scientific Information, National Academy of Sciences—National Research Council, Washington, D.C. Part V, 137.

Oettinger, A. G. (1960). Automatic Language Translation (Harvard).

Bielfeldt, H. H. (1958). Rückläufiges Wörterbuch der Russischen Sprache der Gegenwart (Berlin).

Unbegaun, B. O. (1957). Russian Grammar (Oxford).

ATTITUDINAL MEANINGS CONVEYED BY INTONATION CONTOURS*

ELIZABETH ULDALL
University of Edinburgh

This paper describes experiments in which Osgood's *semantic differential* was used to measure the attitude of listeners to a variety of intonation patterns. 16 pitch contours were applied by synthesis to recordings of four sentences and listeners were asked to rate the patterns with respect to 10 scales of the type BORED/INTERESTED, POLITE/RUDE. From the results it was possible to draw some conclusions about the relative effectiveness of the chosen scales and about some general features of the intonation patterns which had particular weight with respect to three factors: Pleasant/Unpleasant, Interest/Lack of Interest and Authoritative/Submissive.

It is clear that some kind of meaning is conveyed by the intonation of connected speech in both tone-languages (Chang, 1958) and non-tone-languages. There is little agreement about the terms in which this meaning is to be described ; every writer on the subject employs an open-ended supply of terms for this purpose. One kind of meaning conveyed is, however, clearly social and emotional rather than referential. Intonation can express social attitudes: speaker to listener—" It wasn't what she said, it was the way she said it ! " ; to subject matter—" Well, don't get in a temper with me ; *I'm* not the Income Tax collector " ; to the world in general—" He sounds so arrogant ", " Don't whine ! " (Allport and Cantril, 1934).

Attitude measurement (Thurston and Chave, 1929 ; Krech and Crutchfield, 1948) seemed a promising technique by which to attempt to find out whether a group of subjects from the same linguistic community would in fact agree on the " meanings " of intonations, and whether some few very general " dimensions of meaning " in the emotional area could be extracted.

Osgood's *semantic differential* (Osgood *et al.,* 1957) was the attitude-measuring technique used in the experiment which will be described here. The part or aspect of meaning with which the semantic differential can and does deal is precisely the emotional one.

* *The work described here was carried out with the benefit of research funds from the Haskins Laboratories, New York, in connection with a grant from the Carnegie Corporation of New York. It is a pleasure to acknowledge my debt to both organizations.*

The factor analysis mentioned in the text was carried out by Dr. Boris Semeonoff of the Department of Psychology of Edinburgh University, who also supervised the calculation of rank difference correlations. I am very grateful for his help and advice.

My colleagues in the Phonetics Department at Edinburgh have been kind enough to discuss this project with me at length, which has clarified the assumptions underlying the experiment; they have also made helpful suggestions regarding further experiments of this kind.

Stated briefly, the experiment to be described here consisted in presenting the same sentence, with differing intonation contours imposed upon it synthetically, to a set of subjects, who rated each sentence-plus-intonation as to whether it conveyed the impression that the speaker was bored or interested, rude or polite, agreeable or disagreeable, and so on down a list of ten paired opposites, the " scales ".

Synthesized speech was essential in order to be quite sure that all features except the intonation remained the same while the intonation varied. A human speaker making such an array of intonations on the same sentence would at the same time make changes in length, stress and tempo. Since these features appear to convey the same kind of emotional information as intonation (Chang, 1958 ; Fairbanks and Hoaglin, 1941 ; Fairbanks and Pronovost, 1939), it was essential to exclude variations in them.

Four sentences were used, to each of which the various contours were applied in turn. The sentences were:

A. He expects to be here on Friday. (Statement)
B. Did all of them come in the morning ? (Yes-or-no question)
C. What time did they leave for Boston ? (Question-word question)
D. Turn right at the next corner. (Command)

The sentences were intended to be as colourless as possible so as to allow the intonation to add as much as possible to their meaning, and so that they would fit into as many situations as possible when combined with different intonations. In the conduct of the experiment no attempt was made to provide a context of situation for any of the sentences ; possibly it would be better to provide a context or set of contexts, but sound film would be necessary to ensure that all subjects were offered the same context as nearly as possible.

Objection was made by some subjects at the end of the experiment that the command was in fact one which was limited to a smaller range of social situations than the other sentences were. This may be the case with commands in general.

Sentences were chosen consisting of alternations of strong and weak syllables, in order that contours whose effects depend on different treatments of strong and weak could be applied to them. Each contained at least three strong syllables and ended in a strong-weak combination—" Friday ", " morning ", " Boston ", " corner ".

These sentences were recorded as spoken by Dr. Alvin Liberman of the Haskins Laboratories. All were spoken on a steadily falling intonation of rather narrow range.

Sixteen intonation contours were synthesized and applied in turn to the four sentences by means of the Intonator, a component of the Voback synthesizer at the Haskins Laboratories (Borst and Cooper, 1957) ; this removed the original intonation and allowed another to be combined with the spoken material by painting patterns on a pitch-control tape.

The pitch range chosen was 75 c.p.s. to 250 c.p.s. 250 c.p.s. is perhaps rather a high limit for a man's voice, but a wide range was desired partly for the purpose of making some " extreme " contours and partly to make it possible to synthesize contours of the same " shape " but placed in distinctively different parts of the range.

All but four of the contours (see Fig. 1) were perfectly " smooth ", that is, the

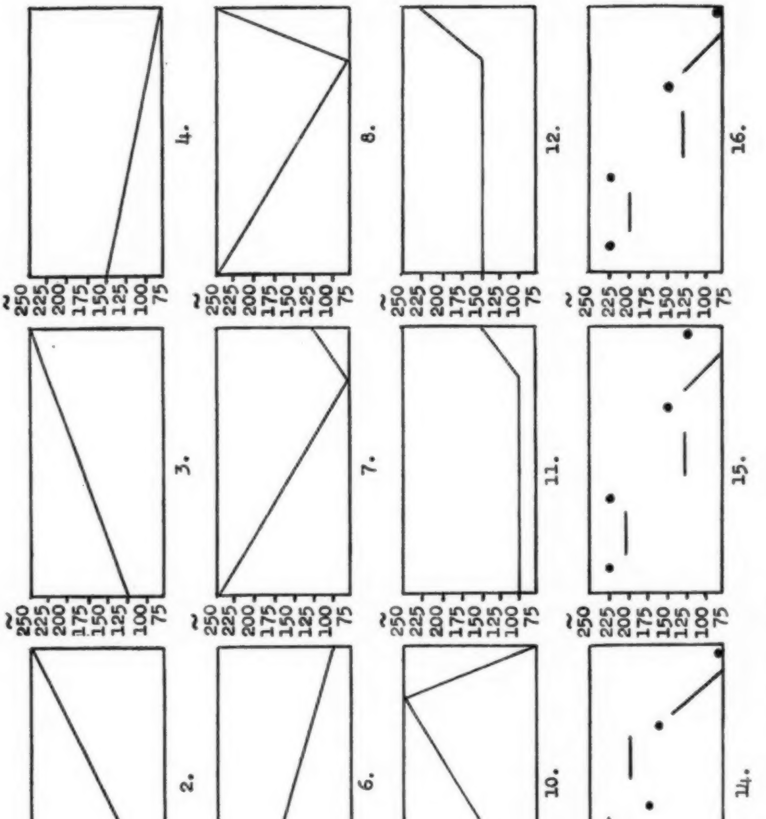


Fig. 1. Intonation contours used in the experiment.

changes in pitch took place at a regular rate in time whatever the sounds, voiced or voiceless, of which the sentence consisted. It was thought that this treatment would make the various sentences on the same contour more readily comparable. These " smooth " contours in fact sound fairly natural. The remaining four contours, shown in a dot-and-dash notation in Fig. 1, nos. 13-16, are contours with " perturbed " weak syllables. These syllables are either, as in contours 15 and 16, on a *higher* pitch than the course of the nearby strong syllables, or, as in contours 13 and 14, on a *lower* pitch. The number of weak syllables between the strong ones is not exactly the same in all the sentences ; the notations in the figure are symbolic of the treatment of weak syllables, not exact maps of any sentence.

The following kinds of difference were incorporated in the contours:

1) Range used. Compare contours 1, 2 and 3 ; 4, 5, and 6 ; 7 and 8 ; 9 and 10 ; 11 and 12.
2) Direction at end. Compare contours 13 and 14 ; 15 and 16.
3) Shape : unidirectional or with a change of direction. Compare contours 5 and 10 ; 2 and 8.
4) Treatment of weak syllables:
   a) Continuing the line of strong syllables: all contours except 13, 14, 15 and 16.
   b) Dropping below the line of strong syllables: 13 and 14.
   c) Rising above the line of strong syllables: 15 and 16.

In all the " smooth " contours in which a change of direction took place, the " turn " was always on the same syllable, the last strong one, so as to reduce the effects of " stress " being differently placed in the different contours. In the contours with " perturbed " weak syllables there was some difference of opinion among listeners as to where the main " stress " fell, though they were intended to give the impression of the main " stress " on the last strong syllable, as in the other patterns.

The " scales " on which the contours were rated consisted of pairs of opposed adjectives. These were chosen after inspecting the results from a pilot experiment using intonations synthesized on the speech synthesizer PAT in the Phonetics Department of the University of Edinburgh (Strevens, 1959). The page to be marked by the subjects for each contour on each sentence was as follows:

BORED — — — — — — — INTERESTED
POLITE — — — — — — — RUDE
TIMID — — — — — — — CONFIDENT
SINCERE — — — — — — — INSINCERE
TENSE — — — — — — — RELAXED
DISAPPROVING — — — — — — — APPROVING
DEFERENTIAL — — — — — — — ARROGANT
IMPATIENT — — — — — — — PATIENT
EMPHATIC — — — — — — — UNEMPHATIC
AGREEABLE — — — — — — — DISAGREEABLE

The terms were arranged with seven places between them ; the subjects were instructed that the places next to the terms should be checked to indicate " extremely " (bored or interested, etc.), the two places a little farther in from the terms to indicate " quite " (bored or interested, etc.), the two places flanking the middle to indicate " slightly " (bored or interested, etc.), and the middle space to indicate " neutral " or " neither " in relation to the scale under consideration.

The assumption made in choosing the " scales ", the pairs of adjectives, was that there are three main kinds of attitude conveyed in intonation: (1) amount or strength of feeling or interest, " emphasis ", (2) pleasantness or unpleasantness of personal relations, (3) a " power " relationship between speaker and listener, authority versus submission.

The twelve subjects in the experiment were seven men and five women, two thirds of them eastern Americans, the rest Americans rather mixed as to education and residence ; they were mostly in their late twenties or thirties. Twelve is rather a small number of subjects for an experiment of this sort, but the subjects gave a satisfactory spread of ratings over most of the scales. This and the high correlations between some of the scales makes it likely that the results have some validity.

The conduct of the experiment was as follows: the subjects first underwent a training period in which they heard one of the sentences (the statement) on all its contours without being asked to write anything. It was thought desirable that they should hear the whole range of variation before being asked to rate any one contour on the scales. For the test itself, which was taken in four parts on four different days, one for each sentence, each intonation contour, on a loop of tape, was played repeatedly while the subjects rated it on a mark sheet as shown above. Ten to fifteen repetitions were generally required before all subjects had given the contour a rating on all of the ten scales. On a signal that all subjects had completed the page referring to that contour, the next contour was played and rated on another page, and so on. The arrangement of the contours was random ; half the subjects were given the test in the original randomized order, the other half in the reverse order.

The seven spaces for scoring, e.g., from " extremely bored " through " quite bored ", " slightly bored ", " neutral ", " slightly interested ", " quite interested ", to " extremely interested ", were given values, 1 to 7. The scores given by the subjects to the various contours on the various scales were averaged ; each contour thus received an average score on each scale. In discussing the results of the experiment any numbers given will be scores of this kind. Thus a score of 2·8 for one contour on the " bored/interested " scale indicates that the average of the scores places it between " quite bored " and " slightly bored ", a little closer to the latter.

Quite a wide scatter of scores over the scales was obtained for the various contours. The subjects obviously found some of the scales easier to use than others in relation to intonation contours. Table 1 shows the range of mean scores for different contours over the various scales. " Bored/interested " in the case of all four sentences was the scale on which the contours varied the most. " Deferential/arrogant ", on the other hand, was clearly a less convincing concept in this connection.

The contours were then arranged in the order of their mean scores on each scale, TABLE 1(a)

A. He expects to be here on Friday.

BORED/INTERESTED	1·5	5·5
TENSE/RELAXED	3·0	5·8
AGREEABLE/DISAGREEABLE	2·9	5·7
POLITE/RUDE	2·7	5·4
SINCERE/INSINCERE	2·0	4·5
DISAPPROVING/APPROVING	2·7	5·0
EMPHATIC/UNEMPHATIC	2·5	4·8
TIMID/CONFIDENT	3·7	5·6
DEFERENTIAL/ARROGANT	3·5	4·8
IMPATIENT/PATIENT	3·3	4·4

B. Did all of them come in the morning ?

BORED/INTERESTED	1·2	5·5
POLITE/RUDE	2·4	5·5
AGREEABLE/DISAGREEABLE	2·5	5·6
DISAPPROVING/APPROVING	2·3	5·1
TENSE/RELAXED	3·5	6·1
SINCERE/INSINCERE	2·5	5·0
EMPHATIC/UNEMPHATIC	2·8	4·9
TIMID/CONFIDENT	3·5	5·2
DEFERENTIAL/ARROGANT	3·1	4·7
IMPATIENT/PATIENT	3·2	4·7

Range of the means for 16 contours on the 10 scales. For each sentence, the scales are ordered with respect to the amount of the scale covered by the mean scores.

e.g., from the " most bored " to the " most interested ". Spearman rank-correlations between these orders were calculated, with corrections for tied ranks. Table 2 shows the correlations between the various word scales for each sentence. Table 3 shows the correlations between the various sentences.

From Table 2 it will be seen that scales 1, 3, 6 and 8 produced predominantly negative correlations. These scales were therefore reflected, i.e., the positions of the defining scale terms were treated as reversed. As a result, i.e., the terms " interested " " confident ", " approving " and " patient " are aligned with the first term in each of the other pairs, and are associated with them in respect of the factor loadings mentioned below.

On the basis of these correlations a factor analysis was made with the object of extracting what may be described as the dimensions of emotional meaning contained in the scales used. The hypothesis in choosing the scales was that an " emphasis " or " interest " dimension would be the first and largest factor to emerge ; however, as in Osgood's experiments, the " evaluative " or " pleasant/unpleasant " factor is by far

TABLE 1(b)

C. What time did they leave for Boston ?

BORED/INTERESTED	2·2	6·0
DISAPPROVING/APPROVING	2·5	5·8
POLITE/RUDE	1·6	4·8
AGREEABLE/DISAGREEABLE	2·1	5·3
SINCERE/INSINCERE	2·4	5·1
TENSE/RELAXED	2·8	5·4
IMPATIENT/PATIENT	2·7	5·3
EMPHATIC/UNEMPHATIC	2·4	4·2
TIMID/CONFIDENT	4·0	5·4
DEFERENTIAL/ARROGANT	3·3	4·5

D. Turn right at the next corner.

BORED/INTERESTED	1·6	6·0
TENSE/RELAXED	2·6	5·8
POLITE/RUDE	2·4	5·5
TIMID/CONFIDENT	2·8	5·8
AGREEABLE/DISAGREEABLE	2·5	5·4
SINCERE/INSINCERE	2·0	4·5
DISAPPROVING/APPROVING	2·8	4·7
IMPATIENT/PATIENT	3·1	5·0
EMPHATIC/UNEMPHATIC	2·5	4·3
DEFERENTIAL/ARROGANT	3·0	4·8

Range of the means for 16 contours on the 10 scales. For each sentence, the scales are ordered with respect to the amount of the scale covered by the mean scores.

the strongest, accounting for more than 50% of the variance, followed by a much less prominent " interest " or " emphasis " factor, about 20%, and " authority/submission " 8 - 13%. It seems reasonable to equate Factor II in this material with Osgood's second factor which he calls " activity ", and Factor III with his third factor which he calls " potency ".

Table 4 shows the results of inspecting the contours, arranged under each word-scale according to mean score, and picking out the tonal-factors common to the most strongly rated contours at each end of the scales, in order to relate them to the meaning-factors extracted by the factor analysis.

I have attempted to get at a possible " conventional neutral " associated with each of the sentence-types by looking for those contours which in each case scored the smallest total difference from the " neutral " score of 4 on all the scales. This procedure does not point out any particular type of contour for each sentence as conveying " least feeling ", but suggests only that, on the whole, contours of small range or small change of direction at the end are rated less strongly than those with more " lively " tonal behaviour.

TABLE 2(a)

A. He expects to be here on Friday.

	BOR.-	POL.-	TIM.-	SIN.-	TEN.-	DIS.-	DEF.-	IMP.-	EMPH.-	AGR.-
	INT.	RUD.	CON.	INS.	REL.	APP.	ARR.	PAT.	UN.	DIS.
BOR.-INT.		−0.52	0.09	−0.17	−0.68	0.42	−0.36	−0.12	−0.37	−0.33
POL.-RUD.	−0.52		−0.57	0.91	0.18	−0.88	0.40	−0.40	0.72	0.83
TIM.-CON.	0.09	−0.57		−0.71	0.20	0.69	0.16	0.37	−0.53	−0.62
SIN.-INS.	−0.17	0.91	−0.71		0.19	−0.86	0.25	−0.34	0.80	0.87
TEN.-REL.	−0.68	0.18	0.20	0.19		0.06	0.17	0.20	0.16	−0.03
DIS.-APP.	0.42	−0.88	0.69	−0.86	0.06		−0.30	0.48	−0.68	−0.83
DEF.-ARR.	−0.36	0.40	0.16	0.25	0.17	−0.30		−0.12	0.01	0.34
IMP.-PAT.	−0.12	−0.40	0.37	−0.34	0.20	0.48	−0.12		−0.05	−0.30
EMPH.-UN.	−0.37	0.72	−0.53	0.80	0.16	−0.68	0.01	−0.05		0.59
AGR.-DIS.	−0.33	0.83	−0.62	0.87	−0.03	−0.83	0.34	−0.30	0.59	

B. Did all of them come in the morning ?

	BOR.-	POL.-	TIM.-	SIN.-	TEN.-	DIS.-	DEF.-	IMP.-	EMPH.-	AGR.-
	INT.	RUD.	CON.	INS.	REL.	APP.	ARR.	PAT.	UN.	DIS.
BOR.-INT.		−0.63	0.23	−0.56	−0.65	0.83	−0.53	0.46	−0.55	−0.71
POL.-RUD.	−0.63		−0.23	0.72	0.59	−0.68	0.25	−0.59	0.53	0.77
TIM.-CON.	0.23	−0.23		−0.37	0.27	0.45	0.17	0.30	−0.65	−0.58
SIN.-INS.	−0.56	0.72	−0.37		0.36	−0.63	0.17	−0.49	0.66	0.72
TEN.-REL.	−0.65	0.59	0.27	0.36		−0.39	0.37	−0.05	0.31	0.37
DIS.-APP.	0.83	−0.68	0.45	−0.63	−0.39		−0.50	0.70	−0.50	−0.87
DEF.-ARR.	−0.53	0.25	0.17	0.17	0.37	−0.50		−0.41	0.03	0.33
IMP.-PAT.	0.46	−0.59	0.30	−0.49	−0.05	0.70	−0.41		−0.10	−0.56
EMPH.-UN.	−0.55	0.53	−0.65	0.66	0.31	−0.50	0.03	−0.10		0.76
AGR.-DIS.	−0.71	0.77	−0.58	0.72	0.37	−0.87	0.33	−0.56	0.76	

Spearman rank-correlation coefficients between the various scales for each sentence.

It is possible that with a larger number of subjects and a treatment of the *modal* score rather than the *mean*, something more like a "conventional neutral" contour for each sentence might emerge.

A study of the results for each of the 16 contours when the scales are grouped in the three categories, Pleasant/Unpleasant, Interest/Lack of Interest, Authoritative/Submissive, shows that certain contours carry particular weight with respect to these factors. No. 6, the narrow-range fall from 200 c.p.s. to 100 c.p.s., was, for instance, the most disliked, and is frequently found rated most strongly unpleasant, on all four sentences. The low narrow-range fall, no. 4, 150 c.p.s. to 75 c.p.s., is often found with it. Narrow range is generally disliked, and "smooth" contours proceeding steadily in one direction (particularly downward) are found less pleasant than "broken" contours with a change of direction or movement up and down of strong and weak

TABLE 2(b)

C. What time did they leave for Boston ?

	BOR.- INT.	POL.- RUD.	TIM.- CON.	SIN.- INS.	TEN.- REL.	DIS.- APP.	DEF.- ARR.	IMP.- PAT.	EMPH.- UN.	AGR.- DIS.
BOR.-INT.		−0.55	0.49	−0.81	−0.60	0.64	−0.61	0.38	−0.83	−0.69
POL.-RUD.	−0.55		−0.07	0.65	0.10	−0.78	0.85	−0.77	0.56	0.87
TIM.-CON.	0.49	−0.07		−0.37	−0.06	0.30	0.03	0.09	−0.45	−0.24
SIN.-INS.	−0.81	0.65	−0.37		0.56	−0.65	0.53	−0.47	0.84	0.73
TEN.-REL.	−0.60	0.10	−0.06	0.56		−0.12	0.11	0.07	0.59	0.23
DIS.-APP.	0.64	−0.78	0.30	−0.65	−0.12		−0.58	0.68	−0.68	−0.89
DEF.-ARR.	−0.61	0.85	0.03	0.53	0.11	−0.58		−0.67	0.37	0.69
IMP.-PAT.	0.38	−0.77	0.09	−0.47	0.07	0.68	−0.67		−0.19	−0.67
EMPH.-UN.	−0.83	0.56	−0.45	0.84	0.59	−0.68	0.37	−0.19		0.74
AGR.-DIS.	−0.69	0.87	−0.24	0.73	0.23	−0.89	0.69	−0.67	0.74	

D. Turn right at the next corner.

	BOR.- INT.	POL.- RUD.	TIM.- CON.	SIN.- INS.	TEN.- REL.	DIS.- APP.	DEF.- ARR.	IMP.- PAT.	EMPH.- UN.	AGR.- DIS.
BOR.-INT.		−0.83	0.15	−0.54	−0.73	0.44	−0.55	0.23	−0.73	−0.65
POL.-RUD.	0.83		−0.33	0.84	0.50	−0.73	0.64	−0.51	0.73	0.89
TIM.-CON.	0.15	−0.33		−0.41	0.37	0.51	0.24	0.19	−0.43	−0.38
SIN.-INS.	−0.54	0.84	−0.41		0.22	−0.94	0.55	−0.76	0.61	0.80
TEN.-REL.	−0.73	0.50	0.37	0.22		−0.10	0.68	−0.05	0.42	0.35
DIS.-APP.	0.44	−0.73	0.51	−0.94	−0.10		−0.51	0.78	−0.52	−0.72
DEF.-ARR.	−0.55	0.64	0.24	0.55	0.68	−0.51		−0.52	0.36	0.52
IMP.-PAT.	0.23	−0.51	0.19	−0.76	−0.05	0.78	−0.52		−0.20	−0.52
EMPH.-UN.	−0.73	0.73	−0.43	0.61	0.42	−0.52	0.36	−0.20		0.59
AGR.-DIS.	−0.65	0.89	−0.38	0.80	0.35	−0.72	0.52	−0.52	0.59	

Spearman rank-correlation coefficients between the various scales for each sentence.

syllables. Weak syllables rising above the surrounding strong ones are on the whole " unpleasant " when followed by a final fall.

It is clear that the various kinds of tonal difference which were incorporated in the sixteen contours convey different meanings on the different sentence types. I had supposed that the emotional effect of a given contour would be more nearly the same on the different sentences than was in fact the case.

In the case of " pleasant " *vs.* " unpleasant ", the value of final rise or fall is different on the statement from what it is on the other sentence types: statements can be " pleasant " while either falling or rising at the end, while on the questions and the command contours with final rises tend to be the " pleasant " ones.

When we look at the factor of " interest " *vs.* lack of interest ", it is the yes-or-no question which is indifferent as to whether the contour is rising or falling finally.

TABLE 3

SENTENCES	BOR.- INT.	POL.- RUD.	TIM.- CON.	SIN.- INS.	TEN.- REL.	DIS.- APP.	DEF.- ARR.	IMP.- PAT.	EMPH.- UN.	AGR.- DIS.
A-B	0.76	0.38	0.17	0.18	0.51	0.15	0.36	−0.18	0.24	0.35
A-C	0.86	0.45	0.39	0.56	0.78	0.73	0.51	−0.11	0.58	0.56
A-D	0.87	0.26	0.69	0.22	0.89	0.40	0.36	0.09	0.33	0.09
B-C	0.68	0.52	0.31	0.49	0.69	0.52	0.14	0.34	0.40	0.58
B-D	0.75	0.81	0.56	0.56	0.66	0.56	0.44	0.51	0.66	0.74
C-D	0.78	0.40	0.46	0.34	0.81	0.60	0.21	0.45	0.47	0.41

Spearman rank-correlation coefficients between sentences for each scale.

The third factor is perhaps too small to discuss profitably, but again a final fall or rise does not seem important in distinguishing the " authoritative " and " submissive " contours ; range takes first place, as it does in the " interest " factor in yes-or-no questions.

The behaviour of unstressed syllables appears to have different values on different sentence types. Rising unstressed syllables are unpleasant and uninterested on the statement and the question-word question, but acceptable on the yes-or-no question if the contour ends in a rise.

These differences must be related to conventional " neutral " or " carrier " tunes for the sentence types in a given speech community.

Further work with larger groups of subjects and other sentences will be necessary to determine whether the differing values of the contours on different sentence types are consistently true of them as types.

It will also be interesting to try similar experiments on groups of speakers of other kinds of English, e.g., R.P. and Scots. Presumably the most nearly usual contours on given sentence types will receive " pleasant " or at least " neutral " ratings ; almost certainly the various contours will be differently ordered by the different speech communities, depending on their own norms. For instance, the fall-plus-low-rise contour, no. 7, was rated by the American group in this experiment, on the yes-or-no question, as " bored, polite, confident, insincere, relaxed, disapproving, arrogant, impatient, emphatic, agreeable ", whereas one would expect it to be rated as more pleasant by speakers of an accent such as R.P. in which it is described as the typical contour for this kind of question.

TABLE 4

	A. Statement		B. Yes-or-no question		C. Question-word question		D. Command	
	Contours		Contours		Contours		Contours	
I.								
PLEASANT	10 4 13 9 14 8	Tendency to wide range Change of direction Lowered weak syllables	13 10 15 9 2 14 8 3	Tendency to wide range Final rise Fall with change of direction	8 14 2 9 10 3 13 1	Wide range Final rise Fall with change of direction Lowered weak syllables	8 1 10 12 9 11 2 15	Tendency to wide range Final rise Final fall with change of direction
UNPLEASANT	6 15 12 5 16	Tendency to narrow range Little or no change of direction Raised weak syllables	4 7 6 16 5	Final fall, especially with narrow range Fall with raised weak syllables	16 5 15 6 4	Narrow range Final fall Raised weak syllables	4 16 6 14 5 7	Final fall, especially without change of direction
II.								
INTEREST	2 12 3 13 1 10 8	Final rise Lowered weak syllables	10 9 5 8 7	Tendency to wide range Change of direction	8 1 2 11 3 14 10 13	Tendency to wide range Final rise Lowered weak syllables	10 8 1 3 11 2 9 12	Final rise Final fall with change of direction
LACK OF INTEREST	6 15 5 7 4 9 16	Tendency to narrow range Tendency to final fall Raised weak syllables	4 6 3 9	Narrow range Little or no change of direction	6 15 4 5 16	Tendency to narrow range Tendency to final fall Raised weak syllables	4 16 6 5 14	Tendency to final fall, especially without change of direction
III.								
AUTHORITATIVE	5 12 2 8 10 9 3	Tendency to wide range	4 6 14	Final fall		No discernible tendencies	4 6 7	Tendency to final fall No perturbed weak syllables
SUBMISSIVE	6 9 7	Tendency to narrow range Perturbed weak syllables	1 2 15	Final rise			3 12 15	Final rise Perturbed weak syllables

References

Allport, G. W. and Cantril, H. (1934). Judging personality from voice. *J. Soc. Psychol.*, 5, 37.

Borst, J. M. and Cooper, F. S. (1957). Speech research devices based on a channel Vocoder. *J. acoust. Soc. Amer.*, **29**, 777.

Chang, N-C.T. (1958). Tone and intonation in the Chengtu dialect (Szechuan, China). ***Phonetica*, 2**, 59.

Fairbanks, G. and Hoaglin, L. W. (1941). An experimental study of the durational characteristics of the voice during the expression of emotion. ***Speech Mono.*, 8**, 85.

Fairbanks, G. and Pronovost, W. (1939). An experimental study of the pitch characteristics of the voice during the expression of emotion. ***Speech Mono.*, 6**, 87.

Krech, D. and Crutchfield, R. S. (1948). Theory and Problems of Social Psychology (New York).

Osgood, C. E., Suci, G. J. and Tannenbaum, P. H. (1957). The Measurement of Meaning (Urbana, Illinois).

Strevens, P. (1959). The performance of PAT. ***Revista do Laboratório de Fonética Experimental, Universidade de Coimbra*, 4**, 5.

Thurstone, L. L. and Chave, E. J. (1929). The Measurement of Attitude (Chicago).

“STRATEGY’’ IN THE SPONTANEOUS UTTERENCE OF NUMBER SYMBOLS

H. B. G. Thomas
University College Hospital, London

This paper discusses the possible significance of different “strategies”, that is of different distribution functions, in the utterance of certain language units. Two samples of the utterance of numbers are analysed and found to exhibit a given strategy. The existence of such strategies has implications with respect to the capacity of the brain for handling both semantic and statistical information.

In a recent study of the distribution of word-length in short complete spontaneous writings (Thomas, 1960) it was shown that certain utterances obeyed one of four distribution-functions with surprising precision. These functions were termed “strategies” and were revealed, in any given sample, as follows: The proportion of words of *S* syllables (p_s) was converted into a quantity termed the *avoidance* (A_s) defined as $-$ log p_s, for each length class. (It will be noticed that avoidance has the dimensions of a *statistical information.*) The avoidances were then plotted against various functions of *S*. A *pure strategy* was said to be obeyed when A_s was found to be linear with respect to some function of *S* over the whole range of *S* employed. The strategies were named according to the nature of this function, f(*S*), viz. linear, logarithmic, log. factorial or quadratic as the case might be.

In order to account for the fact that at least four different pure strategies are met with in different utterances and that the same writer may show different strategies on separate occasions, it was suggested that the coding of concepts into language involved the handling of *information* (carried by patterns of nervous activity) in a set of coding networks, and that the strategy of an utterance was determined by whichever one of these networks was rate-limiting for the handling of the necessary information. It was further postulated that changes in the physico-chemical state of the brain were able to shift the rate-limiting rôle from one network to another.

The same definition of avoidance can be applied to any set of categories and in the present study the author has found that spontaneous and unconstrained utterances containing numerical symbols sometimes show the phenomenon of strategy. In particular, the avoidance of numerical symbols of value *x* (written A_x) has been found to be proportional to log *x*, or to show strong correlation with log *x*, in three samples of which two are presented below.

It is suggested that when a sample of freely uttered numerical symbols obeys any strategy of the general form: $A_x = k'.\mathrm{f}(x)$, where *k*' is a constant, the rate-limiting operation is that in which the neural correlates of numerical concepts are activated, so that “thinking of the numbers” rather than “uttering the number symbols” is the slowest component of the overall process.

The theory implies that the strategy $A_x = k'.\mathrm{f}(x)$ occurs when the “generation-time” of the numerical symbol *X* is determined by the time taken to discriminate the concept " x " from other numerical concepts ; whereas A_x will be independent of x (the usual finding) when the rate-limiting factor is the discrimination and utterance of the symbol itself. This view leads indirectly to a prediction concerning the average spacing of the various number symbols vis-à-vis their avoidances, in numerical utterances which show the strategy phenomenon ; and the prediction is substantially borne out in the one instance where it is tested.

A somewhat special definition of *semantic information* is adopted, which makes it possible to imagine the amount of intensity of " meaning " as a measurable quantity. Thus in the special case of numerical meaning, the concepts " 1 ", " 2 ", " 3 ", ... etc. are held to entail the handling in the brain of successively and systematically increasing amounts of (numerical) semantic information. It is argued that a strategy of the type $A_x = k'.\mathrm{f}(x)$ reveals a functional connection in the brain between the semantic information and the statistical information (i.e., the avoidance) of the number symbols.

Relationships of this kind are interpreted from a philosophical standpoint which holds that the meaning of a word or other symbol resides in the semantic rules and constraints which govern the contexts in which it may be correctly uttered. The fact that the strategy $A_x = k'.\log x$ is discernible in the usage of number-words in a piece of writing by Gertrude Stein, is tentatively explained by supposing that she was using number-words according to more or less fixed semantic rules—that is, with meaning, not randomly, despite the oddity of the prose style. On the other hand, it is suggested that the strategy may occur even when the numbers are uttered " randomly " and free from semantic constraints, if certain types of cerebral dysfunction are present.

Case 1

An elderly man admitted to hospital in one of a series of attacks of confused behaviour. It was not possible to obtain a history of his illness from him except for a complaint about his vision which he could not elaborate.

He was hypertensive, B.P. 220/110, with moderate retinal arteriosclerosis ; the only other material signs found were a mild spastic weakness of the right arm and bilateral extensor plantar responses. There was no visual or tactile inattention, no astereognosis and no apparent disorder of graphical or verbal expression. He was docile and co-operative but his performance on several pencil-and-paper tests was inconsistent and bizarre and at times he seemed merely to be amusing himself. Thus he drew an adequate bicycle ; he drew an arrow pointing in the required direction, without error or delay ; he drew a symmetrical daisy-head and an excellent freehand circle ; yet when asked to put numbers into the latter so as to represent a clock-face, he began at 12 and proceeded correctly until 6, after which he crowded all the remaining numbers into the space normally occupied by 7, 8 & 9. His sketch map of England was a triangle with rounded corners ; France, the Irish Sea, North and South were all correctly indicated but the Isle of Wight was spontaneously placed up on the " West Coast " and the label " North America " was placed due North of England. The labels East and West were interchanged.

The overall impression was of patchy arteriosclerotic brain-damage, producing an

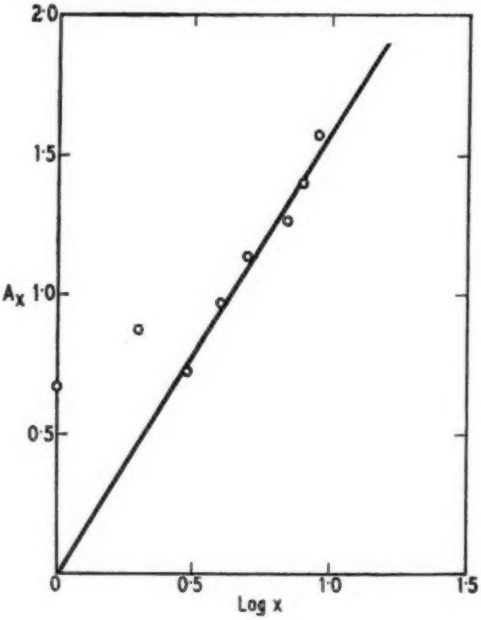


Fig. 1. Six out of nine points obey the law $A_x = 1{\cdot}60 \log x$, in the second utterance of Case 1.

early dementia affecting thinking more than communication. An E.E.G. report read: " . . . generalized dysrhythmia . . . without any focal features . . . compatible with cerebro-vascular disease."

As part of a current investigation, he was asked to fill in, square by square and line by line, from top left to bottom right, a grid of 150 spaces, using the digits 0 - 9 inclusive, in any order or combination he chose. This he did, rapidly and with no interruption, until all 150 were written. There was nothing remarkable about the statistics of this utterance ; the digits were all employed with approximately equal frequency, which is the finding in the great majority of such utterances. However, he was next asked to fill in another grid, omitting completely any one digit he chose. He omitted the digit 6. The result is reproduced below ; and in Fig. 1 it will be seen that the avoidance A_x of the digit x obeyed the equation: $A_x = 1.60 \log x$ for $x = 3, 4, 5, 7, 8$ & 9. The points $x = 1, 2$ did not appear to obey any law ; the point $x = 0$ cannot be plotted, since log 0 is meaningless. However the close adherence of six out of nine points to the same law suggests that this law was latent in his total performance but obscured by other factors over the range $x = 0, 1, 2$.

3	1	9	2	1	8	0	3	2	1
4	3	1	2	0	5	4	3	0	2
2	1	0	3	5	3	4	2	3	1
0	3	0	1	5	0	7	5	4	3
3	4	5	4	3	2	1	5	3	1
1	1	1	0	0	0	1	3	2	1
3	4	5	4	3	2	1	0	3	4
3	0	4	3	2	1	0	8	9	0
9	0	1	3	5	7	9	0	1	3
8	1	2	4	7	8	0	1	2	3
7	0	1	3	8	7	1	2	3	4
5	1	0	2	7	0	2	1	2	4
4	0	1	3	7	1	3	2	1	5
5	0	2	3	7	0	1	2	1	4
4	0	1	3	8	1	2	1	3	3

TABLE 1

x	n_x	A_x	$\log x$	m_x
0	24	0·796	–	5
1	32	0·671	0	3·8
2	20	0·875	0·30	6·5
3	29	0·714	0·48	4·5
4	16	0·972	0·60	7·7
5	11	1·135	0·70	10·5
6	–	–	–	–
7	8	1·273	$0{\cdot}84_5$	13
8	6	1·398	0·90	26·8
9	4	1·574	0·95	27

Data from Case 1.

Case 2

The second example is of a rather different kind. A short story by Gertrude Stein, " Jenny, Helen, Hannah, Paul and Peter ", was analysed in its entirety with respect to the relative frequencies of the seven number-words which occur in it. The story is some 90 pages long and approximately one word in twenty is a number-word. The total number of such words is 1935. The story was chosen for analysis because of this phenomenal preoccupation with numerical words.

As will be seen from Fig. 2, when the avoidance of number-words of value x (A_x) is plotted against log x, the seven points appear to lie evenly scattered about a straight line through the origin. Assuming that a relation of the form $A_x = k' \log x$ is the correct interpretation of the data, and that the scatter is due to interference by irrelevant factors (such as the fact that the number-words were being uttered at the same time as, and intermixed with, other words) the slope of the best-fitting line may be computed by the method of least squares. The result of this procedure is a value of 2·81 for the constant k'.

Discussion

The word " information " is used with a variety of meanings, but these may for most purposes be reduced to two: those of statistical and semantic information. The relationship between these two concepts is one of the major practical and logical problems in the application of information theory to psychology, for they seem to belong to opposite sides of the brain-mind barrier and tend to be thought of as behaviourist and subjectivist ideas respectively.

Statistical information is a measurable quantity, originally defined for inanimate systems by communication engineers but applicable also to human language behaviour. Any channel of communication may be regarded in general as handling M different

TABLE 2

NUMBER-WORD	NUMERICAL VALUE x	n_x	A_x	$\log x$
One	1	1603	0·082	0
Two	2	136	1·153	0·30
Three	3	34	1·755	0·48
Four	4	87	1·347	0·60
Five	5	72	1·429	0·70
Ten	10	1	3·287	1·00
Sixteen	16	2	2·986	1·20

Data from Case 2.

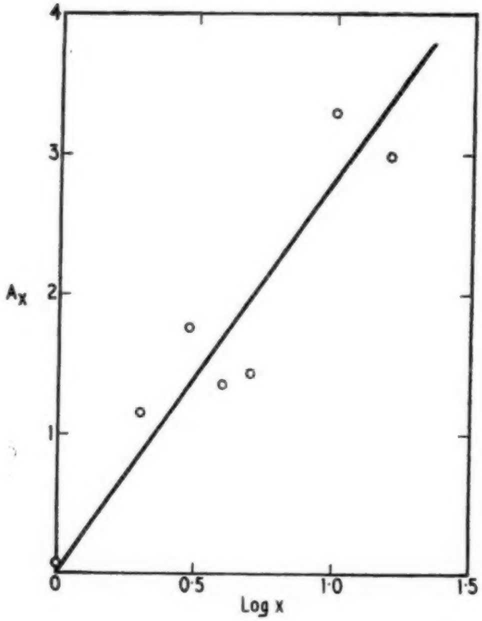


Fig. 2. Points are scattered symmetrically about the line $A_x = 2{\cdot}81 \log x$. Story by Gertrude Stein.

types of signal, each having its own average frequency, or statistical probability of occurrence. If the probability of a particular signal-type J is p_j, the statistical information of each occurrence of a J is defined as $-\log p_j$; it will be noted that this is a quantity with the same formal dimensions as have been assigned above to avoidance.

Now a sample of language may be broken up into arbitrary units, such as words, or sentences, or syllables, and these units may then be classified into any set of categories we choose. For example, the *words* in an utterance might be classified according to their spelling, or the number of syllables they contain, or any other criterion we may decide upon. A quantity of the form $-\log p_i$ can then be computed for each of the M different categories and gives a measure of their statistical information relative to the scale defined by the set of categories employed. Thus it is proper to speak of the information of a particular word-type (or any other signal) only with reference to a stated or implied set of alternative signals, each with a known or predictable probability.

The second usage of " information " is essentially subjective. Semantic information is conveyed by language in the sense that language is able to evoke " meaning " in the mind of a recipient. In order to estimate semantic information quantitatively, it would be necessary first to distinguish and to identify each subjective experience which constitutes a distinct class of meaning ; and secondly, within each of these meaning-classes (quantity, number, time, etc.) it would have to be possible to denote intensities of meaning on a numerical scale.

Let it be supposed then that we can contrive to isolate one such meaning-class for analysis, that the " meaning " in question can be experienced with varying degrees of intensity, and that the degrees of intensity can be assigned numbers so as to define a numerical meaning-intensity scale. In communication between individuals, the degrees of this scale will be represented by a conventional set of symbols. By counting the frequencies with which these representational symbols occur in a given utterance, we might compute the statistical information of each ; this might then be compared with the meaning-intensity number corresponding to each symbol. In this manner the relation between the statistical and the semantic informations could be investigated, for this one series of symbols and meanings, in the given utterance.

Such a programme is perfectly feasible in the case of numerical language. The numerical symbol X (e.g., 3, 8, etc.) stands for the x^{th} degree on the subjective scale of numerical meaning-intensity. It seems reasonable therefore to state as a convention that the semantic information associated with the symbol X must be, if not simply x, then some function of x. In an utterance containing numerical symbols, let us now suppose that the avoidance (statistical information) of X relative to other numerical symbols is found to be A_x. If it happens that a law exists (for this utterance) such that A_x is a function of x, then by the above convention the implication is that a systematic relation exists between the semantic and the statistical information—*in the brain of the subject at the time of the utterance.* That is, the " meaning-intensity " of the concept is determining the relative frequency of utterance and, therefore, the avoidance, of the symbol.

Now it is highly probable that the intellectual functions of the brain depend upon the manipulation of complex spatio-temporal patterns of nerve-impulses which, by virtue of their specificity, convey and embody information. Here we are dealing with a statistical information, which would in principle be computed by counting the frequencies of nerve-impulses passing along different pathways. To do this is not, of course, a practical possibility and the next step involves making an assumption: it is that the amount of information handled by a certain neuronal network in generating any linguistic unit J is proportional to the avoidance (A_J) of that unit in the utterance considered. This is the same as to assume that the statistical information at the neuronal level determines the statistical information at the level of utterance, since avoidance is also a statistical information. We shall adopt this assumption, with the rider that the relevant neuronal network may not be the same one at all times. It leads to a prediction which can be tested:

In the course of a finite utterance, finite numbers of each symbol-type will be produced and we may think of each symbol-type as being uttered at an average repetition-rate, or conversely, as requiring a particular average *generation-time* in the brain. If in fact the avoidance A_x is proportional to the average information handled by the crucial neurones in generating the symbol X each time that it is uttered ; and if, secondly, the brain is analogous to other communication devices in that the time taken to handle a message is proportional to its information content ; then the average time (t_x) taken to generate an X should be linear with respect to A_x.

Assuming that the mean rate of *writing* symbols of all types considered (e.g., the digits 0 to 9) is the same, the average generation-time for any given symbol-type can be estimated by averaging the number of other symbols intervening between successive repetitions of the type in question. Thus, writing m_x for the average number of other digits uttered between successive X's, we can plot A_x against m_x ; if a linear relation between m_x and A_x is found, this result will be compatible with the view that the generation-time fixes the avoidance and indirect support will be given to the interpretation set out above.

In Table 1 the last column shows the values of m_x observed for the utterance of Case 1 and in Fig. 3 it will be seen that an excellent linear relation of the form: $m_x = c.A_x - c'$ (where c, c' are constants) is obeyed for seven of the nine values of x, viz., $x = 0,1,2,3,4,5,7$. The two points corresponding to $x = 8,9$ do not obey the law ; m_8 and m_9 are approximately equal and it seems as though the digits 8 and 9 were exempt from this law just as those at the other end of the numerical range (viz., 0,1 and 2) were exempt from the strategy $A_x = k'\log x$. With this reservation therefore, the test supports the hypothesis that the avoidance, the generation-time and the neuronal information concerned in producing a given symbol-type are all linearly interrelated in, at least, this instance of a " strategic " numerical utterance. (It must be remembered that the subject was required to fill up a prescribed " grid " with digits and that if he had been left free to choose for himself the total number of digits uttered, the aforementioned exceptions might not have occurred.)

In the analysis of the syllabic strategies previously discovered in samples of written verbal language, it was postulated that the various different strategy functions arose because different operations involved in the overall generation of the utterance might become rate-limiting in different utterances. A strictly analogous explanation will account for the two different types of numerical utterance so far recognized. The majority of such utterances show no relation between A_x and x ; the avoidances of the

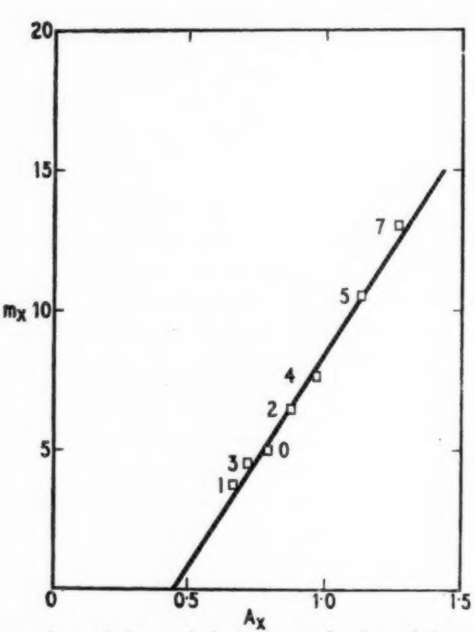


Fig. 3. The average spacings of the symbols (m_x) as related to their avoidances (A_x) in the second utterance of Case 1.

different symbol-types are all approximately equal. Only a small proportion of samples show the phenomenon in which A_x is linear with respect to log x. It may be imagined that the process of uttering a sequence of numerical symbols (digits or words) involves two main intellectual operations—"thinking of" the numerical concept and "thinking of" the appropriate symbol. Should circumstances be such as to make either one of these tasks consistently more time-consuming than the other, the time taken to generate each symbol will be determined by the rate of the slower operation, which would then be termed rate-limiting. According to the discussion above, the character of this operation, by determining the generation-time of each symbol-type, would also determine its avoidance.

Considering the symbols 0 to 9 merely as visual or psychomotor patterns, it is difficult to detect any consistent relation between the complexity of the pattern and the numerical value of the digit; similarly for the corresponding numerical words. There is little to choose between them in point of complexity and we should expect the neuronal information handled by the brain when the various *symbols* are borne in mind to be approximately equal for all. If therefore the rate-limiting step happens to be that of "thinking of the symbol" and if the hypothesis is correct, the avoidances should tend to be equal for all the symbols and A_x should be independent of x. The hypothesis suggests, in short, that in the majority of numerical utterances where the subject is writing numbers " at random ", the rate-limiting process is the expressive phase, in which the number concepts are put into symbolic form and written down.

Conversely, in those rare utterances where a systematic relation is found between A_x and a function of x, the inference seems to be that the numerical value, or " meaning ", of the digit or number-word has determined its avoidance ; and here the hypothesis suggests that the rate-limiting operation is that of " thinking of meaningful numerical concepts ", rather than that of matching the correct symbols to them. In other terms, the semantic information of the numerical concept is determining the statistical information of the corresponding symbol, because the brain is relatively less quick at *handling numerical ideas* than at *formulating numerical symbols*. Clearly, when a subject is ostensibly producing numbers " at random ", very little intellectual effort is needed to choose the numerical concepts to be uttered and we should expect the expressive phase—the " thinking of " and the writing or speaking of the *symbols*—to be rate-limiting in normal circumstances. This expectation is in accord with the fact that A_x is found to be independent of x in most cases. If, however, the numbers uttered were determined as the outcome of some kind of calculation, involving a relatively elaborate manipulation of meaningful numerical concepts, then the conceptual phase might well become rate-limiting in normal people.

According to certain philosophers (Ryle, 1957) the " meaning " of a word or symbol resides in the nature of the rules which conventionally govern its usage ; and words convey meaning from one person to another by virtue of the fact that both individuals know these rules. This approach is especially convincing when applied to numbers and mathematical symbols of all kinds ; it is a familiar fact that in order to be able to perform useful and valid arithmetical operations with numbers (or algebraic operations with letters) it is necessary first to learn and then to obey consistently the rules governing their use. If the rules are not obeyed the results " come out wrong " and are without meaning or reality, for the parallelism between physical operations on physical entities and mental operations on symbolic entities has been broken. On this view, then, the meaningful use of symbols entails the imposition of semantic constraints by the brain ; and since the symbols are no longer being uttered randomly their relative frequencies and the relative probabilities with which one symbol follows another will no longer tend to be equal. We may note that an individual may adopt a self-imposed set of semantic constraints quite consistently, so that his symbolic utterance is strictly meaningful ; but unless other people know the system of constraints, they will not find it so.

In this connection, the utterance of Case 2 is especially interesting. In this instance the numerical symbols were words, not uttered " randomly " but integrated into a discourse largely made up of non-numerical words. Bizarre though Miss Stein's style may be, there is little doubt that she was imposing semantic constraints upon the number-words she employed. The situation resembles the hypothetical case mentioned earlier, in which numbers might be uttered not " randomly " but subject to certain systematic constraints, as in calculation. It may be, then, that the strategy $A_x = k'\log \mathrm{x}$ was manifest in her writing because the conceptual phase was rate-limiting.

Consider now the cases in which A_x has been found to be proportional to, correlated with or linear with respect to log x in ostensibly “ random ” number utterances. Out of more than seventy samples, only two others have shown a tendency similar to that of Case 1. The first of these, produced by a dysphasic right-handed patient with left carotid artery thrombosis, obeyed an equation of the form $A_x = k'\log x + A_0$, but A_0 was large and k' was small so that the line was almost horizontal. The second, produced by a dysphasic man with pronounced hypertensive brain-damage, who could only count perseveratively from “ one ” and reached at each attempt a point where he broke off with an exasperated expletive (“ One, two—damn ! One, two three, four—dammit ! One, two . . . ” etc.) obeyed with some scatter an equation of the form $A_x = k'\log x$. It appears that *brain damage is probably an essential condition for this strategy to be produced in “ random ” utterances of numbers.* Clearly, any lesion affecting the neurones in which the imposition of semantic constraints is carried on could make the conceptual phase rate-limiting, especially under conditions which impose a conceptual load: in accordance with this idea it will be recalled that Case 1 only showed the strategy when under the burden of having to remember not to use any sixes.

In conclusion it should perhaps be pointed out that this evidence drawn from the special case of numerical utterances, whether “ random ” and unconstrained or subject to some degree of semantic control, has a direct bearing on the more general question of “ strategy ” in verbal language as a whole. In that field there is some evidence to support the view that one of the principal syllabic strategies (the logarithmic) is produced under conditions where, once more, the weakest link in the overall process of writing language seems to be that of discriminative thinking, rather than that of expressing thought in words. If indeed the conceptual phase is rate-limiting when such utterances are produced, it is not perhaps surprising that the logarithmic syllabic strategy has so far been found only in certain emotive poems and in the writings of certain active untreated schizophrenics. In addition, the fact that Case 1 produced his first utterance according to a pattern whereby A_x was independent of x, and then produced a second utterance showing the strategy $A_x = k'\log x$ when forbidden to write any sixes, argues that the phenomenon of “ strategy-switch ”, first seen with syllabic strategies in written prose, can occur also in numerical utterance ; and it indicates one kind of circumstance—increased “ conceptual load ”—which may be able to precipitate a “ switch ”.

I would like to thank Dr. E. A. Blake Pritchard and the other physicians at the National Hospital, Maida Vale, whose patients wrote the majority of the number-utterances, for their encouragement ; those friends and acquaintances who wrote the rest, for theirs ; and Dr. D. A. Pond for his constructive criticism and interest.

References

Ryle, G. (1957). British Philosophy in the Mid-Century (London), 239 et seq.

Stein, G. (1951). Two: Gertrude Stein and her Brother (New Haven), 143 et seq.

Thomas, H. B. G. (1960). Syllabic “ strategies ” in written language. *Nature*, **185**, 485.

INDEX TO VOLUME 3, 1960

AUTHORS

GENERAL

Papers